Universidad de Alcalá
COMISIÓN DE ESTUDIOS OFICIALES
DE POSGRADO Y DOCTORADO

ACTA DE EVALUACIÓN DE LA TESIS DOCTORAL

Año académico 2016/17

DOCTORANDO: **AHMADZADEH, AMIR MASOUD**
D.N.I./PASAPORTE: ******1608**

PROGRAMA DE DOCTORADO: **D347 DOCTORADO EN TECNOLOGÍAS DE LA INFORMACIÓN Y LAS COMUNICACIONES**
DEPARTAMENTO DE: **TEORÍA DE LA SEÑAL Y COMUNICACIONES**
TITULACIÓN DE DOCTOR EN: **DOCTOR/A POR LA UNIVERSIDAD DE ALCALÁ**

En el día de hoy 08/06/17, reunido el tribunal de evaluación nombrado por la Comisión de Estudios Oficiales de Posgrado y Doctorado de la Universidad y constituido por los miembros que suscriben la presente Acta, el aspirante defendió su Tesis Doctoral, elaborada bajo la dirección de **JOSÉ LUIS PORTILLA FIGUERAS // LUCAS CUADRA RODRÍGUEZ.**

Sobre el siguiente tema: *CONTRIBUTION TO THE DESIGN AND OPERATION OF ADVANCED MOBILE COMMUNICATIONS SYSTEMS*

Finalizada la defensa y discusión de la tesis, el tribunal acordó otorgar la CALIFICACIÓN GLOBAL[1] de (**no apto, aprobado, notable** y **sobresaliente**): Sobresaliente

Alcalá de Henares, 8 de Junio de 2017

EL PRESIDENTE

Fdo.: David Camacho

EL SECRETARIO

Fdo.: Sancho Salcedo

EL VOCAL

Fdo.: Carlos Casanova

Con fecha 29 de junio de 2017 la Comisión Delegada de la Comisión de Estudios Oficiales de Posgrado, a la vista de los votos emitidos de manera anónima por el tribunal que ha juzgado la tesis, resuelve:

- [x] Conceder la Mención de "Cum Laude"
- [ ] No conceder la Mención de "Cum Laude"

La Secretaria de la Comisión Delegada

FIRMA DEL ALUMNO,

Fdo.: Amir M. Ahmadzadeh

[1] La calificación podrá ser "**no apto**" "**aprobado**" "**notable**" y "**sobresaliente**". El tribunal podrá otorgar la mención de "cum laude" si la calificación global es de sobresaliente y se emite en tal sentido el voto secreto positivo por unanimidad.

INCIDENCIAS / OBSERVACIONES:

1. Tanto el presidente Titular, como el vocal Titular, no han podido asistir al acto de defensa por razones personales, habiéndose conformado el tribunal con el Presidente Suplente, Dr. David Camacho Fernández, y vocal suplente, Dr. Carlos Casanova Mateo.

2. El nombre del Director de la Tesis es Dr. José Antonio Portilla Figueras, y no "José Luis Portilla Figueras", como erróneamente se indica en este acta de evaluación.

Universidad de Alcalá

COMISIÓN DE ESTUDIOS OFICIALES
DE POSGRADO Y DOCTORADO

En aplicación del art. 14.7 del RD. 99/2011 y el art. 14 del Reglamento de Elaboración, Autorización y Defensa de la Tesis Doctoral, la Comisión Delegada de la Comisión de Estudios Oficiales de Posgrado y Doctorado, en sesión pública de fecha 29 de junio, procedió al escrutinio de los votos emitidos por los miembros del tribunal de la tesis defendida por *AHMADZADEH, AMIR MASOUD*, el día 8 de junio de 2017, titulada *CONTRIBUTION TO THE DESIGN AND OPERATION OF ADVANCED MOBILE COMMUNICATIONS SYSTEMS*, para determinar, si a la misma, se le concede la mención "cum laude", arrojando como resultado el voto favorable de todos los miembros del tribunal.

Por lo tanto, la Comisión de Estudios Oficiales de Posgrado resuelve otorgar a dicha tesis la

MENCIÓN "CUM LAUDE"

Alcalá de Henares, 11 julio de 2017
EL PRESIDENTE DE LA COMISIÓN DE ESTUDIOS
OFICIALES DE POSGRADO Y DOCTORADO

Firmado digitalmente por VELASCO PEREZ JUAN RAMON - DNI 03087239H
Fecha: 2017.07.12 15:32:20 +02'00'

Juan Ramón Velasco Pérez

Copia por e-mail a:
Doctorando: AHMADZADEH, AMIR MASOUD
Secretario del Tribunal: SANCHO SALCEDO SANZ.
Directores de Tesis: JOSÉ LUIS PORTILLA FIGUERAS // LUCAS CUADRA RODRÍGUEZ

Universidad de Alcalá

Campus Universitario
Dpto. de Teoría de la Señal y Comunicaciones
Ctra. Madrid-Barcelona, Km. 36.6
28805 Alcalá de Henares (Madrid)
Telf: +34 91 885 88 99
Fax: +34 91 885 66 99

Dr. D. JOSÉ ANTONIO PORTILLA FIGUERAS y Dr. D. LUCAS CUADRA RODRÍGUEZ, Profesores Titulares de Universidad del Área de Conocimiento de Teoría de la Señal y Comunicaciones de la Universidad de Alcalá,

CERTIFICAN

Que la tesis "**Contribution to the Design and Operation of Advanced Mobile Communications Systems**", presentada por D. Amir Masoud Ahmadzadeh, realizada en el Departamento de Teoría de la Señal y Comunicaciones bajo nuestra dirección, reúne méritos suficientes para optar al grado de Doctor, por lo que puede procederse a su depósito y lectura.

Alcalá de Henares, 11 de enero de 2017.

Fdo.: Dr. D. José Antonio Portilla Figueras

Fdo.: Dr. D. Lucas Cuadra Rodríguez

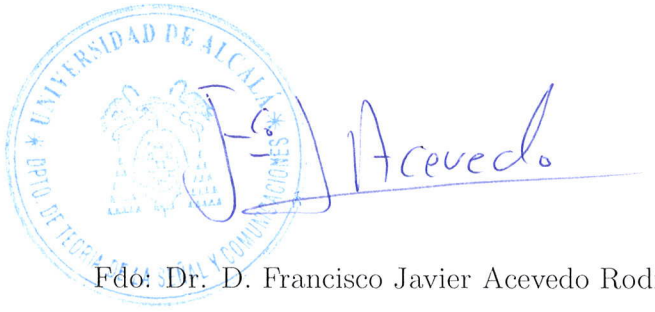
Universidad de Alcalá

Campus Universitario
Dpto. de Teoría de la Señal y Comunicaciones
Ctra. Madrid-Barcelona, Km. 36.6
28805 Alcalá de Henares (Madrid)
Telf: +34 91 885 88 99
Fax: +34 91 885 66 99

D. Amir Masoud Ahmadzadeh ha realizado, en el Departamento de Teoría de la Señal y Comunicaciones, y bajo la dirección de los doctores D. José Antonio Portilla Figueras y D. Lucas Cuadra Rodríguez, la tesis doctoral titulada "**Contribution to the Design and Operation of Advanced Mobile Communications Systems**", cumpliéndose todos los requisitos para la tramitación que conduce a su posterior lectura.

Alcalá de Henares, 11 de enero de 2017.

EL DIRECTOR DEL DEPARTAMENTO



Fdo: Dr. D. Francisco Javier Acevedo Rodríguez.

Universidad de Alcalá

Escuela Politécnica Superior

Departamento de Teoría de la Señal y Comunicaciones

Programa de Doctorado en
Tecnologías de la Información y las Comunicaciones (D445)

Doctorate Thesis

Contribution to the Design and Operation of Advanced Mobile Communications Systems

Author:

Amir Masoud Ahmadzadeh

Thesis supervisors:

José Antonio Portilla Figueras, Ph.D.

Lucas Cuadra Rodríguez, Ph.D.

2017

Acknowledgment

I would like to express my deepest appreciations to my supervisor professor, Dr. José Antonio Portilla Figueras, for his great and persistent direction and guidance, that made this dissertation possible. He is and will always be a role model for me, who taught me the hard work, efficient thinking and the willingness to help.

I thank and share the credit of my work with Dr. Lucas Cuadra Rodríguez, for his support and dedication to complete this thesis and our research publications.

I thank University of Alcalá, the international coordinators and the post-graduate office, for helping me follow my interests in studies. I also thank the director and members of GHEODE research group at Signal Theory and Communication department for creating a professional and friendly research environment.

Lastly, the words alone cannot express the thanks I owe to my life partner Cristina, for her love and constant support, and to my parents, that have always been there for me.

Amir M. Ahmadzadeh

Resumen

Aunque el conjunto de tecnologías 3G (Third Generation) basadas en WCDMA (*Wide-band Code Division Multiple Access*) y HSPA (*High-Speed Packet Access*) forman en la actualidad el fundamento de las redes móviles de banda ancha más empleadas, LTE (*Long-Term Evolution*) es, sin embargo, la tecnología móvil con el crecimiento *más rápido* de toda la historia. LTE es capaz de proveer velocidades de transmisión de datos *muy elevadas* y con una *latencia extremadamente reducida*, especialmente en el enlace descendente, DL (*downlink*), a expensas de incrementar su *complejidad*. Éstas son algunas de las características que convierten a LTE en una tecnología de banda ancha móvil con enormes posibilidades, no solo para usuarios domésticos sino también para aplicaciones novedosas en comunicaciones M2M (*machine-to-machine*), servicios de salud tipo *mHealth*, o los nuevos servicios asociados a las ciudades inteligentes (*smart grids*). Algunas cifras que ilustran este vertiginoso crecimiento es que el número de líneas LTE ha aumentado con una tasa muy elevada en el primer cuatrimestre de 2016, aportando aproximadamente 150 millones de nuevos usuarios, y alcanzando un total de 1200 millones de clientes en todo el mundo. Este desarrollo se aprecia mejor si se considera dentro de un *contexto global* en el que se observa cómo, por ejemplo, las redes basadas en WCDMA han crecido sólo en 70 millones de usuarios, mientras que, incluso, aquellas basadas en GSM (*Global System for Mobile Communications*) han disminuido sus clientes en unos 60 millones –a pesar de que la mayor parte de los usuarios de sistemas 3G y 4G (*Fourth Generation*) utilizan GSM y EDGE (*Enhanced Data rates for GSM Evolution*) como redes de respaldo–.

Hay varias razones que explican el rápido despliegue de las redes LTE. Por un lado, el mercado de dispositivos móviles está creciente a tasas muy elevadas ($\approx 80\%$), constituyendo ya los teléfonos inteligentes (*smartphones*) el 75% de los dispositivos de usuario. Por otro lado, la técnica OFDMA (*Orthogonal Frequency Division Multiple Access*) en el enlace descendente permite que LTE sea capaz de soportar tasas de datos más elevadas que las de HSPA. Hay incluso nuevos comportamientos de usuarios, sobre todo del segmento de población más joven, que están modificando sus preferencias a la hora de ver TV, desplazándose progresivamente desde el televisor convencional al streaming en sus smartphones. La combinación de todos estos factores está espoleando una demanda fuertemente creciente de servicios novedosos y muy intensivos en Mbps (HDTV, vídeo bajo demanda (VoG), gaming, etc.): el tráfico de datos ha crecido un $\approx 65\%$ entre los primeros cuatrimestre de 2015 y el

de 2016.

LTE y *Mobile Worldwide Interoperability for Microwave Access* (Mobile WiMAX), han constituido las dos tecnologías de banda ancha que se han propuesto como respuesta a la iniciativa IMT (*International Mobile Telecommunications*)-Advanced. Cuando comenzamos esta tesis en 2012 no estaba claro aún cuál de estas dos tecnologías en competencia sería la vencedora. En la actualidad, LTE se ha erigido como la vencedora indiscutible en el amplio campo de aplicación de las redes celulares, mientras que Mobile WiMAX ha quedado reducido a nichos de mercado más específicos (aunque enormemente importantes) como la industria de la aviación, smart grids, o ciertos sistemas de comunicación para compañías petrolíferas.

Además de la utilización de OFDM, LTE y Mobile-WiMAX tiene otros aspectos en común, como la posible integración de técnicas AMC (*Adaptive modulation and coding*) y MIMO (*Multiple-Input and Multiple-Output*) para poder proveer conexiones más fiables y rápidas. Sin embargo, y a pesar de estas *similitudes*, LTE y Mobile-WiMAX constituyen soluciones tecnológicas muy *complejas* que exhiben numerosas *diferencias*. En particular, un elemento crucial que aumenta la dificultan tecnológica inherente a ambas se encuentra se encuentra en su flexibilidad en el sentido de que se dejan abiertos muchos aspectos de configuración para que cada compañía pueda diseñar la opción que considere más adecuada. El rendimiento de estas redes depende pues de la forma en la que se soluciona el problema de optimización entre la adecuación de los recursos disponibles en cada configuración a los recursos consumidos por los usuarios.

Dentro de este contexto, el objetivo de esta tesis es estudiar la *capacidad* definida como "el número máximo de usuarios *simultáneos*, con *diferentes perfiles de servicios*, que puede soportar cada nodo de acceso, para cada *configuración del sistema*". Para poder calcular de forma adecuada la capacidad y la tasa binaria (*throughput*) se requiere calcular de forma dinámica la *cantidad de recursos disponibles*, que, a su vez, depende de cómo se forma la correspondiente MAU (*minimum allocation unit*) tanto en LTE como en WiMAX. Ambas tecnologías también tienen en común que el cómputo de éstos es complejo debido a que parte de los recursos (*overhead*) tiene que utilizarse para labores necesarias de señalización, control o sincronización.

La validez de los modelos obtenidos se ha probado de forma exitosa en una gran variedad de configuraciones posibles en ambos tecnologías.

Extended Abstract

Mobile technologies, which involve not only networks but also devices, services and applications, are currently one of the main pillars of our globalized economy, mobile access *networks* being one of the fastest growing industries of all times. Specifically, cellular networks are now evolving to broadband networks, helping users access to their desired data (and to the associated services and applications) at any time and any place. This is shaping our life style quickly and deeply, and generating a new need, "the need to be always connected". The amount of digitally generated data is increasing every day, along with services that require high transmission rates. This is causing a very strong pressure on mobile network operators since users increasingly demand higher and higher bit rates. In response, and benefiting from liberalized markets (but at the same time, pressured by an increasing competition), technology provider companies are adopting constantly novel advances in both cellular networks and devices in the effort of improving users' experience. This strongly growing market is one of the driving forces that is compelling mobile operators to adapt and upgrade their systems and network infrastructures aiming at providing faster access, while accommodating higher system availability and coverage. As mobile cellular networks evolve, the technologies they are based on, and the processes of design, planning and implementation of networks get more and more complex and costly. Unlike traditional cellular networks with only one single user profile (voice-only), the deployment plans of novel generation mobile networks need to ensure that the systems are well configured to provide a variety of services with different requirements.

The vertiginous evolution that mobile communications have undergone in the last decade –basically because of the appearance of new access mobile technologies, novel services and high-tech user devices– has contributed to reach the aforementioned determining position in global economy. There are several access technologies for providing broadband services over wireless networks and, in particular, over mobile networks. There exist currently 6440 million subscriptions worldwide using 3rd Generation Partnership Project (3GPP) mobile networks: Global System for Mobile Communications (GSM) networks, Wide-band Code Division Multiple Access (WCDMA) networks (also known Third Generation (3G) networks), and Forth Generation (4G) cellular networks, such as Long Term Evolution (LTE). In the effort of providing the best customer experience, mobile operators combine these networks to form a mobile access ecosystem of heterogeneous networks. This term is often used

to name a global network that consists of several mobile network technologies (GSM, WCDMA and LTE) covering the same geographical area, and also to describe the coexistence of different non-homogenous cells (macrocells, microcells, femtocells) to guarantee high data rates in small areas with large densities of users.

While WCDMA/High-Speed Packet Access (HSPA) technology is currently the most used cellular technology, Long Term Evolution (LTE) is the *fastest-growing mobile technology*. LTE is able to provide unprecedented *very high data rate* and *extremely low latency*, specially in downlink (DL) at the expense of increasing complexity. High data rate makes LTE very useful not only for conventional users (either via the LTE cellular network itself or by offloading part of traffic to user's WiFi network) but also for novel applications in machine-to-machine (M2M) communication, mHealth services, smart grids, or green LTE. This is one of the reasons why LTE subscriptions have grown at a high rate during Q1 2016, with 150 million new subscriptions during this quarter, reaching a total of 1.2 billion worldwide. During Q3 2015 LTE had added around 120 million novel subscribers, while WCDMA had added only 70 million users. Global System for Mobile Communications (GSM) have even declined by 60 million –in spite of most users of 3G and 4G systems use GSM and Enhanced Data Rates for GSM Evolution (EDGE) networks as a fallback–.

There are several causes explaining this vertiginous development. On one hand, user equipment (UE) manufacturers have made public an annual growth in 2015 of about 79%, smartphones being 75% of all users' devices and the strongest driving force for the aforementioned fast deployment. On the other hand, Orthogonal Frequency Division Multiple Access (OFDMA) technique in DL makes LTE able to support higher data rates than those of HSPA. Thus, LTE *acceptance* (up to 442 commercial LTE networks in 147 countries) is driven by a strong demand for both better user experience and faster networks than those provided by HSPA technology. There are also new users' behaviors, especially those of teens, who are changing the TV/video viewing conduct, moving from TVs to streaming video *on smartphones*. All these combined factors are fueling an increasing the demand of novel, more data-intensive services (such as TV, Video on Demand (VoD), gaming, etc.): data traffic has grown 65% between Q1 2015 and Q1 2016.

Among all the solutions currently *in the market*, 4G mobile networks are the most advanced standardized and interoperable solutions for provision of high-speed mobile data access. LTE and Mobile Worldwide Interoperability for Microwave Access (Mobile WiMAX) are the two state-of-art sets of mobile broadband technologies. When we began this thesis in 2012, Mobile WiMAX and LTE was the two main technologies competing for the International Mobile Telecommunications (IMT)-Advanced initiative, and it was not clear at that moment which of the two technologies would win and lead the Broadband Mobile Internet. However, at present, LTE is the undisputed winner for broadband mobile cellular networks, while WiMAX has resulted in being confined to very specific market niches such as aviation industry, smart grids, smart city, point-to-point long distance or high rate radio transmission in applications for oil and gas companies.

Both LTE and Mobile WiMAX technologies use the Orthogonal Frequency Division Multiplexing (OFDM) technique, which is an efficient resource sharing method to map the users' data in time and frequency space. They both can integrate enhancing techniques such as Adaptive Modulation and Coding (AMC) and Multiple Input Multiple Output (MIMO) to provide higher data-rates and more reliable connections. LTE and Mobile-WiMAX are however very *complex* solutions that, apart from technological similarities, have many structural and implementation differences. Additionally, the standards that define these two broadband mobile approaches contain numerous flexibilities. The *configuration choices* are left open for the vendors to find their optimal solution according to their service area. The performance of these networks is therefore highly dependent on solving the complex design problem of finding the optimized match between the available system resources and its load. Network planning and optimum design in novel generation mobile networks are active research line. High-level and, at the same time, accurate performance modeling are of high value in planning cellular networks, as full scale dynamic simulations are not affordable in time because of hard complexities for large planning scenarios.

With this in mind, the main purpose of this thesis is to provide a Performance Key Indicator (KPI) for 4G broadband mobile technologies, namely *capacity*, defined as "the maximum number of simultaneous multi-profile users that are jointly supported with each specific access point, under specific system configuration". The key point to accurately estimate the *capacity* and user's *data throughput* is that they require to dynamically compute the *available* resources and the *minimum allocation unit* (MAU) formation in LTE and Mobile WiMAX systems. Both technologies have in common the fact that computing properly these MAUs is a very difficult tasks since part of such resources have to be reserved for a number of necessary network functionalities such as *signaling*, *control* or *synchronization* ("overhead", $\mathcal{O}$), so that the *useful* user's data throughput must be computed as the *available* resources minus those used for overhead. Aiming at carrying out this within a *unified research framework*, we have considered in this thesis the following requirements. Specifically:

1. **System Configuration.** We have studied in detail the technological features of LTE and Mobile-WiMAX, considering the different system configuration options. We have modeled the *available resources* for each configuration and *resource formation and allocation* types:

    - As the number of users increases, the operational system overheads to keep track of these users also increase. So, apart from the fixed system overheads and data load consumed by the users, a dynamic overhead calculation is required to know the remaining resources at each moment.
    - The resource allocation mechanisms of the technology under study needs to be mathematically formulated to obtain the available resources under each specific configuration and to know to what extent each technology shares these available resources between its users.

2. **Service Profile.** Those users located in the service area may demand different services, have different signal reception conditions, and consume different amount of system resources. Thus it is necessary for an accurate modeling to elaborate a "Service Profile" definition to model the user's condition and demand.

3. **System Capacity.** An interactive algorithm is required to compare the system *available resources* to the *demanded resources* aiming at determining whether or not the resources can meet the new demands when other extra user tries to enter the system. Regarding this, we have proposed a *novel algorithm to compute the capacity* in both LTE and Mobile WiMAX systems. It consists of two processes that works in parallel, and estimate, respectively, the amount of real available resources (once those used for PHY+MAC overhead have been removed, using a *system configuration* and a *user distribution profile*) and the amount of resource consumed, using a *service profile*. The algorithm ends in providing the capacity as the maximum number of supported users (N) fulfilling the condition that the available resources are greater than the demanded ones.

With these ideas in mind, for the two main technologies competing for the IMT-Advanced initiative (LTE and Mobile-WiMAX), the *twofold purpose* of this thesis consists in:

1. **Accurately modeling the useful data throughput.** This requires to compute in parallel the *available resources* –removing those used for *overhead*– and the *demanded resources* by several multi-service users).

2. **Accurately modeling the capacity,** defined as the maximum number of simultaneous multi-service users that a cell is able to support with a given Quality of Service. As mentioned, capacity is the KPI we have used as a representation of the network operational capacity and corresponds to the *goodput* (real throughput once overhead has been removed) of the system at full-load.

We have carried out a number of numerical experiments aiming at 1) computing the useful data throughput and the maximum number of users under a variety of different possible configuration sets, and 2) discussing their practical implications.

The proposed models for both useful data throughput and capacity can be used as the basis for more accurate design, planning and dimensioning of 4G networks, as well as to better estimate the cost of implementation.

Contents

Resumen **i**

Extended Abstract **iii**

List of Figures **ix**

List of Tables **xv**

List of Acronyms **xvii**

1 Motivation and related work **1**

1.1 Introduction 1

1.2 Motivation: are mobile operators dying from their success? 4

1.3 Thesis purpose 5

1.4 Related work 12

1.5 Thesis structure 16

2 Exploring the Maximum Number of Simultaneous Multi-service Users in Mobile WiMAX **19**

2.1 Introduction 20

2.2 Technical Background 22

2.2.1 An overview of Mobile WiMAX frame 23

2.2.2 The Time Domain in the Frame Structure 24

2.2.3 Frequency Domain and Subchannelization in Frame Structure 25

2.2.4 Data Mapping 26

2.2.5 Resource Allocation 27

2.2.6 Application Classes and Quality of Service 29

2.3 Related Work 30

2.4 Overview of the Proposed Capacity Estimation Algorithm 32
2.5 Dynamic System Resource Calculation 35
2.5.1 A model to estimate PHY+MAC overhead the number os useful data symbols (N_{sym}) 35
2.5.2 Modulation and Coding Schemes Distribution 37
2.6 Service delivery Model and Subscribers Resource Consumption 39
2.6.1 Data-rate requirements 40
2.6.2 Application profile and number of active users 41
2.6.3 Subscribers Service Demand Calculation 43
2.7 Experimental work 44
2.7.1 Experiment I 45
2.7.2 Experiment II 48
2.8 Summary and conclusions 50

3 Overhead, Data Throughput and Capacity in LTE Downlink 51
3.1 Introduction 51
3.2 Related work 54
3.3 Problem statement 57
3.3.1 Throughput and overhead: mathematical approach 57
3.3.2 Additional background 61
3.4 Overhead mechanisms in Physical Layer for Downlink 67
3.4.1 Overhead mechanisms with period 1 TTI 67
3.4.2 Overhead mechanisms with period 1 frame 73
3.5 Total overhead $\mathcal{O}_{\text{TOT}}^{\Delta t}$ 74
3.6 Overhead and Initial Controlling Symbols 75
3.7 Overhead and maximum number of UEs (with identical N_{CCE} value) per TTI 76
3.8 Realistic scenario: multiple users with different services 77
3.9 Experimental work 78
3.9.1 Reference case 78
3.9.2 Exploring the influence of the Hybrid ARQ configuration . . . 84
3.9.3 Influence of MAT and resource allocation types 85
3.9.4 Maximum number of UE per TTI, $N_{\text{UE}}^{\text{TTI}_i}|_{\text{MAX}}$ 87
3.9.5 Multiple users with different services 90

3.10 Summary and conclusions 91

4 Conclusions and Future Work **95**

4.1 Summary and conclusions 95

4.2 Future work 100

List of Publications **103**

References **105**

List of Figures

1.1 (a) Multiple spatial layer concept. (b) Time-frequency resource structure (normal cyclic prefix case) and MAU. (c) Example of overhead. See the main test for further details. 7

1.2 Simplified representation of the OFDMA-TDD frame structure in Mobile WiMAX. The minimal allocation unit in this example is a "slot" (one sub-channel in frequency dimension and two symbols in the time dimension –in the PUSC permutation mode–). 8

1.3 Flowchart of the proposed algorithm to compute system capacity. It consists of two processes that works in parallel, and estimate, respectively, the amount of real available resources (once those used for PHY+MAC overhead have been removed, using a *system configuration* and a *service profile*) and the amount of resource consumed. n is an internal variable that accounts for the number of users. The algorithm ends in providing the *capacity* as the maximum number of supported users, N. 11

2.1 Simplified representation of the OFDMA-TDD Frame structure in Mobile WiMAX. TTG and RTG stand for "Transmit Transition Gap" and "Receive Transition Gap", respectively. 23

2.2 OFDMA-TDD Frame structure indicating control messages and resource allocation procedure. See the main text for further details. . . 28

2.3 Proposed Capacity Estimation Algorithm. It is formed by two processes that works in parallel, and estimates, respectively, the amount of real available resources (once those used for PHY+MAC overhead have been removed) and the amount of resource consumed. n is an internal variable that accounts for the number of users. The algorithm ends in providing the maximum number of supported active users, N. 34

2.4 Simple representation of the bandwidth partitioning. Based on this data-rate requirement classification, the total available bandwidth is divided into *guaranteed* and *non-guaranteed* partitions. See the main text for further details. 40

2.5 Dynamic trend of the algorithm in Experiment I: DL optimal goodput (green curve) and minimum service-load (red curve), both in kbps, as a function of the number of users to be served. 46

2.6 (a) Number of overhead symbols as a function of the number of active users to be served. (b) Number of overhead unallocated slots as a function of the number of active users to be served. 47

2.7 Resource consumption and available resources (kbps), in both "Case 1" and "Case 2", as a function of the number of supported users (Experiment II). See the main text for further details. 49

3.1 (a) Multiple spatial layer concept. (b) Time-frequency resource structure (normal cyclic prefix case). (c) Example of overhead. See the main text for further details. 53

3.2 Simplified representation of an example of the Resource Allocation Types when BW= 5 MHz. See the main text for further details. . . . 63

3.3 Simplified representation of Physical Channel Processing. See the main text for further details. 64

3.4 Data and Controling Symbols in Downlink TTI visualization, representing the PDCCH, PHICH, PCFICH, and RS signals, and considering the REG and CCE grouping conceptrs 68

3.5 Summary of steps to compute $\mathcal{O}_{\text{PDCCH}}^{\text{TTI, N}_{\text{UE}}^{\text{i}}}$. 72

3.6 Total overhead $\mathcal{O}_{\text{TOT}}^{\Delta t}$ (RE) as a function of BW (MHz). Blue $\square$-symbols represent $\mathcal{O}_{\text{TOT}}^{\text{frame}}$, while red $\bigcirc$-symbols show the overhead at TTI scale, $\mathcal{O}_{\text{TOT}}^{\text{TTI}}$. 80

3.7 $\Delta\mathcal{O}_{\text{TOT}}^{\Delta t}(\%)$ as a function of the bandwidth BW (MHz). Blue square symbols ($\square$) represent $\Delta\mathcal{O}_{\text{TOT}}^{\text{frame}}$, while red circular symbols ($\bigcirc$) represent $\Delta\mathcal{O}_{\text{TOT}}^{\text{TTI}}$. Black $\Diamond$ symbol and $\bigstar$ correspond to those overhead estimated in [72] and [66], respectively. 80

3.8 (a) Percentage of each overhead component $\Delta\mathcal{O}_{\text{comp }i}^{\text{TTI}}(\%)$ with respect to the total overhead at TTI scale $\mathcal{O}_{\text{TOTAL}}^{\text{TTI}}$, as a function of BW (MHz). (b) Percentage of each overhead component $\Delta\mathcal{O}_{\text{comp }i}^{\text{frame}}(\%)$ with respect to the total overhead at frame scale $\mathcal{O}_{\text{TOTAL}}^{\text{frame}}$, as a function of BW (MHz). The inset defines the symbols used for any component. 81

3.9 (a) Data throughput $\mathcal{T}$ (Mbps) as a function of BW (MHz) computed by using the parameter values listed in Table 3.10 and the overhead values shown in Figure 3.7. Red columns represent the throughput computed during 1 TTI, $\mathcal{T}_{\text{TOT}}^{\text{TTI}}$, while blue columns represent the throughput computed during 1 frame, $\mathcal{T}_{\text{TOT}}^{\text{frame}}$. (b) Relative error in throughput estimation, $\varepsilon_{\mathcal{T}_{\text{TOT}}}(\%)$ (Expression 3.46), as a function of BW (MHz). 83

3.10 (a) Throughput (Mbps) as a function of BW (MHz) parametrized by $N_g = 1/6, 1/2, 1, 2$, respectively. (b) Throughput relative error $\varepsilon_{N_g}(\%)$ defined in Expression 3.47. 85

3.11 Lower bound ($N_g = 2$) of the peak data rate (Mbps) as a function of BW (MHz), parametrized by n and q: 1) SISO, with $n = 1$ and $q = 0$. 2) 2×2 MIMO with $q = 0$. 3) 2×2 MIMO with $q = 1$. 4) 4×4 MIMO with $q = 0$. 5) 4×4 MIMO with $q = 1$. 87

3.12 Overhead percentage (%), with respect to the total REs, as a function of BW, for $n = 4$ (a), $n = 2$ (b), and $n = 1$ (c). 88

3.13 Maximum number of users per TTI, parametrized by $N_g = 1/6, 1/2, 1, 2$, as a function of BW for different configurations: SISO (a); 2×2 MIMO with $q = 0$ (b) and $q = 1$ (c); 4×4 MIMO with $q = 0$ (d) and $q = 1$ (e). 89

3.14 (a) Throughput $\mathcal{T}$ (Mbps) of each of the ten users whose DL conditions have been listed in Table 3.12. Blue columns represent each user's $\mathcal{T}_j$ when $w_j = 0.1$, while red columns show $\mathcal{T}_j$ when coefficients w_j are adjusted aiming at equalize each users' $\mathcal{T}_j$ (mean value = 6.927336 Mbps, variance = 998 bps). (b) Corresponding number of RBs that each user UE_j receives during TTI_i, $N_{\mathrm{RB}}^{\mathrm{UE}_j}$ (see Expression (3.6)) when coefficients w_j are adjusted aiming at equalizing $\mathcal{T}_j$. . . . 91

List of Tables

2.1 List of acronyms used in this chapter. 22

2.2 Summary of SOFDMA PHY parameters in PUSC mode as a function of the bandwidth BW. SC means subcarrier. 26

2.3 Block-size (bits) as a function of the Modulation and Coding Scheme (MCS) in PUSC (Partially Used SubCarrier) permutation mode. . . 27

2.4 Relation of each QoS data-rate specification with its corresponding application and service flow category in Mobile WiMAX [147]. 41

2.5 Service Delivery Model. Value of the different parameters used in the applied profiles as a function of the application class. 43

2.6 Input Parameters of Experiment I. The "system parameters" are the default values recommended by WiMAX Forum system evaluation methodology and also are the common values used in practice. The "subscription service parameters" are based on the statistical data, as explained in the presented application profile. 45

2.7 Estimated capacity for different system parameters. The subscriber service parameters are the same as those listed in Table 2.6 and MCS distribution is the one in Table 2.3. "Gua." stands for guaranteed. . 49

3.1 List of acronyms used in this paper. 55

3.2 Effective Code Rate values used in LTE (labeled $\mathrm{ECR}^{\mathrm{MCS}}$) as a function of the Channel Quality Indicator (CQI). MCS stands for Modulation and Coding Scheme. Q is the modulation order. Information extracted from [149]. 59

3.3 Number of resource blocks (N_{RB}) and Resource Block Groups size (P) as a function of LTE bandwidth (BW) [157]. 61

3.4 Summary of layer mapping cases for spatial multiplexing in a configuration with up to 4×4 MIMO, derived from [159]. It provides factor $\Omega_{\max}$ as a function of the layer mapping used, depending on the number of layers N. 65

3.5 Downlink Resource Allocation Types (Type 0, Type 1, and Type 2), and the description of the DCI formats f_{DCI}. Open-loop MIMO does not require knowledge of the channel at the transmitter. Derived from [158]. 67

3.6 PDCCH Construction Parameters [159]: N_{CCE}, N_{REG}, and PDCCH Format as a function of the L_{PDCCH} interval. See [14] for further details. 70

3.7 Expressions to compute the DCI length (L_{DCI} [bits]) as a function of the downlink scheduling type [14]. The values of a, b, and c, used for indicating the number of transmission antennae are listed in Table 3.8. Table 3.5 describes the DCI formats. 70

3.8 Values of parameters a, b, and c to be used in Table 3.7. 70

3.9 Number of PHICH groups as a function of BW and N_g. Information obtained from 3GPP TS 36.211 [162]. 72

3.10 Values of the parameters that minimize the total overhead in the reference study case. See Table 3.5 for further details. 79

3.11 Comparison of LTE maximum data-rate in DL with highest fixed values for: bandwidth (BW = 20 MHz); MCS order (64 QAM with $Q = 1$); deployed spacial multiplexing (MIMO 4×4), and variable values for overhead, ECR, Ω, and Δt. 84

3.12 Values of the parameters involved in the multi-user scenario simulated. ID means identifier. MAT stands for Multiple Antenna Technology. 90

List of Acronyms

2G: Second Generation

3G: Third Generation

3GPP: 3rd Generation Partnership Project

4G: Forth Generation

5G: Fifth Generation

AMC: Adaptive Modulation and Coding

ARPU: Average Revenue Per User

ARQ: Automatic Repeat Request

BE: Best Effort

BLER: Block Error Rate

BS: Base Station

BW: Bandwidth

C-RNTI: Cell's Radio Network Temporary Identifier

CAGR: Compound Annual Growth Rate

CAPEX: Capital Expenditure

CBR: Constant Bit Rate

CC: Convolutional Coding

CDMA: Code Division Multiple Access

CID: Channel Identifier

CQI: Channel Quality Indicator

CR: Contention Ratio

CRB: Contiguously allocated Resource Blocks

CRC: Cyclic Redundancy Check

CTC: Convolutional Turbo Coding

DCD: Downlink Channel Descriptor

DCI: Downlink Control Information

DL: Downlink

DU: Dense Urban clutter

ECR: Effective Code Rate

EDGE: Enhanced Data Rates for GSM Evolution

eNB: Evolved Node B

ErtPS: Extended Real-time Polling Service

FDD: Frequency Division Duplex

FEC: Forward Error Correction

GDP: Gross Domestic Product

Gpbs: Gigabits per second

GSA: Global Supplier Association

GSM: Global System for Mobile Communications

HARQ: Hybrid Automatic Repeat Request

HSDPA: High-Speed Downlink Packet Access

HSPA: High Speed Packet Access

HSUPA: High Speed Uplink Packet Access

ICT: Information and Communications Technology

IE: Information Element

IMS: IP Multimedia Subsystem

IMT: International Mobile Telecommunications (IMT)-Advanced initiative

IoT: Internet of Things

IP: Internet Protocol

ISP: Internet Service Provider

ITU: International Telecommunication Union

LTE: Long-Term Evolution

M2M: machine-to-machine

MAC: Media Access Control sub-level

MAT: Multiple Antenna Technology

MAU: minimum allocation unit

MB: MegaByte

MBB: Mobile Broadband

Mbps: Megabits per second

MCS: Modulation and Coding Scheme

MHz: Megahertz

MIMO: Multiple Input Multiple Output

MS: Mobile Station

MSS: mobile subscriber stations

MTC: Machine-Type Communication

NFV: Network Function Virtualization

nrtPS: Non Real-time Polling Service

OFDMA: Orthogonal Frequency-Division Multiple Access

OPEX: Operational Expenditure

OS: Operating System

PCFICH: Physical Control Format Indicator Channel

PDCCH: Physical Downlink Control Channel

PDSCH: Physical Downlink Shared Channels

PDU: Protocol Data Unit

PHICH: Physical Hybrid-ARQ Indicator Channel

PHY: Physical level

PLM: Propagation Loss Margin

PM: Propagation model

PMP: Point to Multiple Point

PRB: Physical Resource Block

PUSC: Partially Used SubCarrier permutation mode

QAM: Quadrature Amplitude Modulation

QoS: Quality of Service

QPSK: Quadrature phase-shift keying

R: Rural clutter

RAN: Radio Access Network

RB: Resource Block

RBG: Resource Block Group

RCS: Rich Communication Services

RE: Resource Element

ROHC: Robust Header Compression

RRC: Radio Resource Control

RRM: Radio Resource Management

RS: Reference Signal

RTG: Receive Transition Gap

RTP: Real-time Transport Protocol

rtPS: Real-time Polling Service

RTT: Round Trip Time

SC: Sub-carrier

SDU: Service Data Unit

SISO: Single-Input and Single-Output

SNIR: Signal to Noise plus Interference Ratio

SOFDMA: Scalable Orthogonal Frequency Division Multiple Access

SS: Synchronization Signal

TB: Transport Block

TBS: Transport Block Size

TCP: Transmission Control Protocol

TD-SCDMA: Time Division-Synchronous Code

TDD: Time Division Duplex

TTG: Transmit Transition Gap

TTI: Transmission Time Interval

TxD: Transmit Diversity

U: Urban clutter

UCD: Uplink Channel Descriptor

UDP: User Datagram Protocol

UE: User Equipment

UGS: Unsolicited Grant Service

UL: Uplink

VBR: Variable Bit Rate

VoD: Video on Demand

VoIP: Voice over IP

VoLTE: Voice over LTE

VRB: Virtual Resource Block

WCDMA: Wideband Code Division Multiple Access

WiMAX: Worldwide interoperability for microwave access

WRC: World Radio-communication Conference

Chapter 1

Motivation and related work

1.1 Introduction

Mobile technologies are undoubtedly one of the fastest growing industries of all times, and are currently one of the main pillars of worldwide economy. The constant evolution that mobile communications have undergone in the last decades –basically because of the appearance of new access cellular technologies, novel services and novel devices– has contributed to reach this position in global economy [1–4]. The mobile industry (networks + devices) is one of the main driving forces of economic growth, entrepreneurship and prosperity worldwide. It currently generates about 4% of global Gross Domestic Product (GDP), and directly employs about 13 million people, 15 million employees being expected by 2020 [2]. The contribution of mobile industry to the world's GDP is expected to grow at a faster rate than that of the rest of the economy, contributing approximately 4.2% by the end of 2020. These figures illustrate the key impact of mobile industry on economic productivity, although does *not* include other wider intangible socio-economic effects in the "Networked Society" [4] (for instance, those deep changes in our life style produced by "the need to be always connected").

Specifically, mobile industry is one of the driving forces of the new "digital economy". It is fueling innovation in online commerce, mobile money services, digital contents, social networks, etc. Such a "mobile ecosystem" is the cornerstone of the novel economy based on mobile broadband networks, advanced tablets, smartphones, laptops, and a growing variety of other connected devices. In this respect, the number of machine-to-machine (M2M) connections is expected to reach 1 billion by 2020 [2]. Both governments and companies are beginning to realize the huge changing potential of the Internet of Things (IoT) [5, 6], with an increasing number of devices, services and applications such as, for example, wearables and smart cities [7–9]. IoT is expected to overcome mobile phones as the largest class of connected devices in 2018 [1]. Between 2015 and 2021, IoT is expected to boost at a Compounded Annual Growth Rate (CAGR) of about 23 %, reaching by 2021 almost 16 billion of the total predicted 28 billion connected devices.

Additionally, from a social viewpoint, mobile technologies have also the potential to put into practice cost effective solutions to a number of social challenges by facilitating the access to healthcare and education. In particular, these are key challenges in developing countries, where large parts of the population are living in inaccessible rural areas, and in which there are limited funding from government to support healthcare and education social services. In this respect, both mobile networks and related services are producing social benefits by improving social cohesion, education, healthcare, and financial inclusion [2, 3].

From a technology viewpoint, the world is currently facing a very quick technology migration to higher mobile broadband networks, smartphones and other connected devices. According to [2], mobile broadband connections will account for about 70% of global mobile connections by 2020, up from just under 40% at the end of 2014. Smartphone adoption is now reaching critical mass in developed markets, accounting for approximately 60% of devices. Emergent developing counties –fueled by the increased non-expensive accessibility of smart devices– are expected to generate most of the future growth, by adding more than 2.9 billion smartphone users by 2020 [2, 3].

Specifically, with regard to networks, there are several access technologies for providing broadband services over cellular networks. According to the Global mobile Suppliers Association (GSA) there are currently over 6450 million subscriptions worldwide using 3rd Generation Partnership Project (3GPP) mobile networks [10]: Global System for Mobile Communications (GSM) networks, Wide-band Code Division Multiple Access (WCDMA) networks (also known Third Generation (3G) networks), and Forth Generation (4G) cellular networks [10–12], such as Long Term Evolution (LTE) [13, 14]. In the effort of providing the best customer experience, mobile operators combine these networks to form a mobile access "ecosystem of Heterogeneous Networks". The term Heterogeneous Network (HetNet) [15] is often used to name a global network that consists of several mobile network technologies (GSM, WCDMA and LTE) covering the same geographical area [16], and also to describe the coexistence of different non-homogenous cells (macrocells, microcells, femtocells [17–19]) to guarantee high data rates in small areas with large densities of users.

Mobile subscriptions worldwide are getting larger around 3% year-on-year (Q1 2016, compared to the value measured exactly a year earlier), and have reached 7.4 billion. In particular, India grew the most during Q1 2016 (+21 million), followed by Indonesia (+5 million), the US (+3 million) and Pakistan (+3 million). Mobile broadband subscriptions –High Speed Packet Access (HSPA), LTE, and Mobile Worldwide Interoperability for Microwave Access (WiMAX)– are rising by around 20 % year-on-year, increasing by approximately 140 million in Q1 2016 alone [1].

Within this mobile ecosystem, HSPA, based on WCDMA, is the most widely used deployed mobile *broadband* technology in the world. WCDMA/HSPA continues to undergo a significant growth worldwide [1]. In fact, HSPA is not a unique technology, but a set of technologies that allow mobile operators to easily *upgrade* their already

deployed WCDMA networks to support a very efficient provision of speech services and mobile broadband data services (high speed Internet access, music-on-demand, and TV and video streaming, to name just a few). Currently, 83% of mobile operators worldwide are investing and upgrading their WCDMA-based 3G networks [20] and have 1830 million WCDMA subscribers [10]. WCDMA/HSPA technology is expected to cover 90 percent of the world's population by 2020, serving about 3.8 billion subscribers [3]. WCDMA/HSPA added around 30 million during Q1 2016 [1]. These figures illustrate the importance of properly planning and dimensioning HSPA networks [21].

While WCDMA/HSPA technology is currently the most used cellular technology, LTE is the fastest-growing mobile technology [3, 22]. LTE is able to provide unprecedented very high data rate and extremely low latency, specially in downlink (DL) at the expense of increasing complexity [23–26]. High data rate makes LTE very useful not only for conventional users [3] (either via the LTE cellular network itself or by offloading part of traffic to user's WiFi network [27]) but also for novel applications in M2M communications [28], mHealth services [29], smart grids [30], or green LTE [31, 32], even when potentially deployed in TV white spaces [33].

LTE subscriptions have grown at a high rate during Q1 2016, with 150 million new subscriptions during this quarter, reaching a total of 1.2 billion worldwide. Smartphones subscriptions go also on increasing, and are expected to overcome those related to basic phones in Q3 2016 [1]. During Q3 2015 LTE had already added around 120 million novel subscribers, while WCDMA [12] had added only 70 million users, and even Global System for Mobile Communications (GSM) had declined by 60 million [3] –in spite of most users of Third and Four Generation (3G, 4G) use GSM and Enhanced Data Rates for GSM Evolution (EDGE) networks as a fallback [3]–.

There are several causes explaining this vertiginous development. On one hand, user equipment (UE) manufacturers have made public an annual growth in 2015 of 79%, smartphones being 75% of all UEs and the strongest driving force for the aforementioned fast deployment. On the other hand, Orthogonal Frequency Division Multiple Access (OFDMA) technique in DL [34, 35] makes LTE able to support higher data rates than those of High-Speed Packet Access (HSPA) [36, 37]. Thus, LTE acceptance (up to 494 commercial LTE networks in 157 countries [1]) is driven by a strong demand for both better user experience and faster networks than those provided by HSPA technology [13]. There are also new users' behaviors, especially those corresponding to teens, who are changing the TV/video viewing behavior, moving from traditional TVs to streaming video on smartphones [1].

All these combined factors are fueling an increasing demand of novel, more data-intensive services (such as TV, Video on Demand (VoD), gaming, video communication, etc.): data traffic has grown 60% between Q1 2015 and Q1 2016 [1]. This growth in mobile data traffic is caused by both the rising number of smartphones (in particular, LTE smartphones) and the rising data consumption per user. Furthermore, according to [1], the CAGR in 2015-2021 is expected to grow annually, by

application category, in 55% (video), 41% (social networking), 37% (software download), 25% (web browsing) and 19% (in file sharing). The strong trend in video growth is strengthened by the increasing use of embedded video in social media and web pages, which are considered as video traffic in this context [1].

The negative counterpart of the superior performance of LTE when compared to other mobile technologies is its highest complexity, which makes much more difficult its modeling, dimensioning [26,38] and planning [39–41], demanding new simulation and evaluation tools [42] along with novel algorithms for resource allocation [43] and handover in LTE macrocell-femtocell networks [44], to name a few.

There is another negative element for mobile operators: users' demand for IP-based communications services such as Skype, WhatsApp or Facebook Messenger are becoming increasingly very popular, but these do *not* provide almost any benefit to operators. These services are expected to gain even more impulse with the growth of LTE networks and novel smart devices. This is one of the reasons why mobile operators worldwide have to consider which type of over-the-top integration models will allow them to *reach revenue* and *sustain* their business models in the long term. This is a key factor because revenues from more conventional services are under pressure, and operators have significant investment obligations as they deploy LTE networks. This along with *competitive and regulatory difficulties* have led to a *slowdown in operators revenue growth.*

1.2 Motivation: are mobile operators dying from their success?

All the technological innovation presented in Section 1.1 has led to the aforementioned set of thriving business related to mobile networks. Specifically, the market penetration of the different mobile technologies and services had been continuously growing since they were massively introduced at the beginning of the 1990's, leading to the corresponding accelerated growth of the revenues of Mobile Network Operators (MNO). The aforementioned technological context described in the previous section is currently influenced by: 1) a strong *competition* between operators, 2) the introduction of over-the-top services that do not generate significant operator's revenues, and 3) some consequences of the 2008 economical crisis, which will be described later on.

Powered by the emergent variety of novel smart devices, applications and services, data traffic is growing vertiginously, and is expected to undergo an almost ten-fold increase by 2020 [2]. However, the revenues of mobile operators are not growing at the same rate as they used to a few years ago. The consequences of the 2008 economic and financial crisis along with both competitive and regulatory difficulties have led to a *deceleration of revenue growth* [45–48]. In particular, the investment on new network infrastructures reduced dramatically [49] in the period 2008-2012. In Spain, investment dropped from 2,439M€ in 2007 to 1,575M€ in 2012

(35% reduction) [50].

These causes have compelled networks operators to start a trend that aims at exploiting the existing network infrastructures as much as possible, in the effort of using the deployed networks in the most efficient way [51], implementing a variety of services such as voice, best effort data, streaming, and multicast techniques [52, 53] on their networks.

Fortunately, after cutting investment during the peak of the financial crisis in 2009 and 2010, capital expenditure (CAPEX) has started to rise in the effort of deploying mobile broadband networks, as a response to the ever increasing users' demand for higher data rates. However, revenue growth is expected to slow further over the coming years, with a CAGR of 3.1%, lower than 4% in the period 2008-2014 [2].

Governments, companies, entrepreneurs, and the whole society can benefit significantly from advanced mobile technologies, but only in the case in which mobile operators are able to upgrade their existing networks and deploy new networks. Mobile operators need to construct a sustainable business model for new investment, which will persuade investors to provide the required funding.

This implies to optimize the dimensioning of the radio access network resources. A proper dimensioning should alining at: 1) accurately computing the data throughput (to satisfy users' demand for ever increasing bit rates), and 2) estimating the maximum number of simultaneous multi-service users that a cell is able to support with a given Quality of Service (QoS). This twofold aim is the main purpose of this thesis.

1.3 Thesis purpose

When we began this thesis in 2012 there were two main technologies competing for the International Mobile Telecommunications (IMT)-Advanced initiative, LTE and WiMAX [54, 55], and it was not clear at that moment which of the two would win and lead the Broadband Mobile Internet [56].

Both LTE and Mobile WiMAX standards use a number of common technologies with subtle differences [57–63]. The main common technology is OFDMA [57]. In WiMAX, OFDMA is used on both the downlink and the uplink, whereas in LTE it is only used on the downlink. The technology used in the uplink of LTE, single-carrier frequency division multiple access (SCFDMA), is nothing but a simple modification of OFDMA [57]. There are some important reasons for using OFDMA such as the scalability of operation in different bandwidths, the ability to reduce multi-path problems, the ability to handle different data rates, and the the easy combination with Multiple Input Multiple Output (MIMO) antenna techniques [60]. Both LTE and Mobile WiMAX can integrate enhancement techniques such as Adaptive Modulation and Coding (AMC) and MIMO to provide higher data-rates and more better communications [54, 60].

LTE and Mobile-WiMAX are *very complex solutions* [57–63] which, apart from the technological similarities mentioned before, have many structural and implementation differences (see [57,60,61]). Additionally, the standards that define these two broadband mobile solutions contain many *configuration flexibilities*. Such configuration choices have been left open for vendors to find their optimal solutions according to their service area. The performance of these networks is therefore highly dependent on solving this complex optimization problem, i.e. system vs. load. The network optimum planning and design in both LTE and Mobile WiMAX networks is an active research line, as will be shown in Section 1.4. High-level and, at the same time, accurate performance modeling are of high value in planning broadband cellular networks since full scale dynamic simulations are not affordable because of their complexity and computational time required for large planning scenarios.

Within this context, Figures 1.1 and 1.2 will assist us in introducing the purpose of this thesis in a more technical way. They aim at illustrating the concepts of *resource formation* and *minimum allocation unit* (MAU) in LTE and Mobile WiMAX.

On one hand, Figure 1.1 consists of three subfigures. Figure 1.1 (a) aims at illustrating the fact that the DL transmission resources in LTE exhibits dimensions of time, frequency and space. This latter, measured in "layers", is used by means of Multiple Antenna Technology (MAT), which involves several multi-antenna transmission and reception schemes (for instant, MIMO). Figure 1.1 (b) focuses only on one of such layers, and represents, in a simplified way, how OFDMA is used in the LTE DL, in the particular case of Frequency Division Multiple Access (FDD) and normal cyclic prefix. As illustrated, the basic unit for data transmission, called Physical Resource Block (PRB) or simply Resource Block (RB), is a two-dimensional (time, frequency) object in a grid. In time dimension, an RB last 0.5 ms, the length of 1 slot. Two slots form an 1-ms length subframe. A subframe is the *smallest time interva*l in which LTE can assign RBs to UEs (or simply, users), and is called Transmission Time Interval (TTI). In the frequency dimension, an RB consists of 12 orthogonal subcarriers (SCs) assigned to an UE during one time slot, which in turn consists of a number of 7 OFDM symbols. The MAU is formed, as illustrated, by 2 RBs in 1 TTI. Furthermore, the smallest information unit is the one carried by 1 SC during 1 time symbol (T_S), and is called Resource Element (RE). Most of the REs –white REs in Figure 1.1 (c)– are used to transmit *user data* on specialized physical channels such as Physical Downlink Shared Channels (PDSCHs). The rest of the REs, a reduced number of them, have to be reserved for a number of necessary network functionalities such as *signaling*, *control* or *synchronization* [30,64–67] at both the Physical (PHY) level and Media Access Control (MAC) sub-level. These REs are "overhead" ($\mathcal{O}$) REs. Although it will be explained in detail later on, controlling the physical channels is motivated, for instance, by the need to decide, depending on user's Channel Quality Indicator (CQI), the Modulation and Coding Scheme (MCS) to be applied on each user's REs. Other overhead REs have to be reserved for Reference Signals (RSs) (dashed grey REs) –which transport some signaling information– and Synchronization Signals (SSs). The later has not been represented in Figure 1.1 (c) since, unlike the others, are *not* computed at TTI scale, but at *frame* scale.

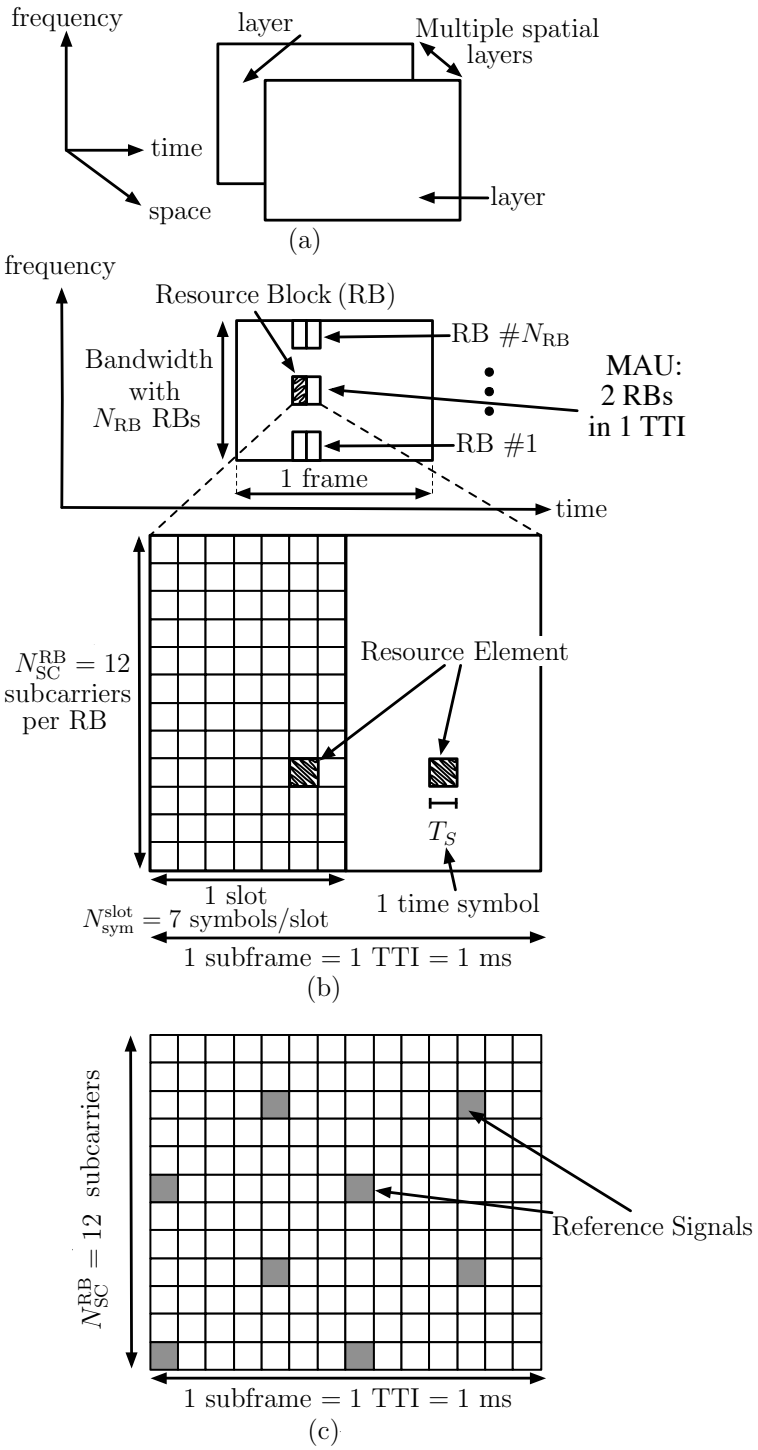


Figure 1.1: (a) Multiple spatial layer concept. (b) Time-frequency resource structure (normal cyclic prefix case) and MAU. (c) Example of overhead. See the main test for further details.

On the other hand, Figure 1.2 will assist us in introducing the MAU in Mobile WiMAX. The standard [68] defines both Time Division Duplex (TDD) and Frequency Division Duplex (FDD) modes for the PHY layer, which uses Orthogonal Frequency Division Multiple Access (OFDMA) technique. In this thesis we have considered TDD-OFDMA since, when we published our work [69], the first releases were expected to use only TDD.

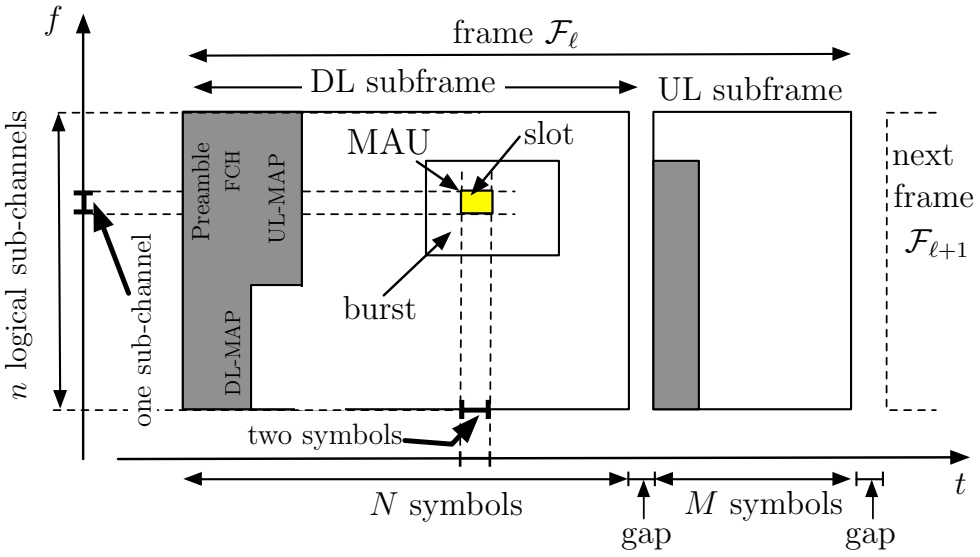


Figure 1.2: Simplified representation of the OFDMA-TDD frame structure in Mobile WiMAX. The minimal allocation unit in this example is a "slot" (one sub-channel in frequency dimension and two symbols in the time dimension –in the PUSC permutation mode–).

Figure 1.2 basically represents a simplified structure of a TDD-OFDMA *frame*, which is a two-dimensional grid in time (t) and frequency (f). In time dimension, the frame consists of a downlink (DL) subframe and uplink (UL) subframe [56]. Each of them transports a number of OFDMA symbols. In frequency dimension, the frame consists of a number of orthogonal subcarriers (SCs) grouped into "sub-channels". This association between physical subcarriers and logical sub-channels is called a "permutation mode". Partially Used Sub-channelization (PUSC) is the most common. In PUSC permutation mode, subcarriers forming a sub-channel are selected randomly from all available subcarriers. The minimal allocation unit in this

case is called a "slot". The size of a slot depends on the permutation mode used. For example, in the PUSC mode a slot is formed by one sub-channel in frequency dimension and two symbols in the time dimension. Most of the frame slots (white areas in Figure 1.2) are used to transmit users' data. Each of these user's *data region* is called "burst" and is a two-dimensional allocation of slots, that is, a group of contiguous sub-channels transporting a set of contiguous OFDMA symbols [70]. However, part of the frame (gray area) must be reserved for the proper system operation (synchronization, signaling, and control). Although it will be explained in detail throughout this thesis, we here briefly review the need form some of these elements. For instance, a "preamble" is used for time *synchronization*, while the "downlink map" (DL-MAP) and "uplink map" (UL-MAP) are used to communicate both the burst-start time and burst-end time, along with the modulation schemes and forward error control (FEC) technique for each mobile station (MS). The other remaining element in Figure 1.2, the "Frame Control Header" (FCH), is used to specify the lengths and subcarriers to be used in these MAPs.

Figures 1.1 and 1.2 show the resource and MAU formations in LTE and Mobile WiMAX, and have in common the fact that calculating properly these MAUs are very difficult tasks to accurately compute the real user data rates. Since part of the resources have to be reserved for a number of necessary network functionalities such as *signaling*, *control* or *synchronization* (overhead, $\mathcal{O}$), the useful user's data throughput must be computed (over 1 TTI) as the useful resources minus those used for overhead.

With these ideas in mind, the twofold purpose of this thesis consists in accurately modeling: 1) *useful data throughput* (which requires to compute in parallel the available resources –removing those used for *overhead*– and the demanded resources by several multi-service users), and 2) *capacity* (the maximum number of simultaneous multi-service users that a cell is able to support with a given Quality of Service). Capacity is a Key Performance Indicators (KPI) we have used as a representation of the network operational capacity and corresponds to the *goodput* (real throughput once overhead has been removed) of the system at full-load. The obtained capacity can be used as the basis for network design, planning and dimensioning purposes, as well as cost estimation of the new generation cellular networks implementation.

As will be shown in the literature review in Section 1.4, overhead is usually estimated either as a constant value (regardless of operation conditions), or as a percentage of the total available resources, or assuming some overhead mechanisms as negligible, leading to overestimate bit rates or throughputs. Another important difficulty encountered when trying to analyze overhead is that most of the related information is spread over a number of recommendations, technical books or research papers, and to the best of our knowledge there is no paper modeling and discussing the influence that *all* the overhead mechanisms have on different aspects of Mobile WiMAX and LTE performance. Our *dynamic* overhead calculation methodology, which involves both concepts of the Physical (PHY) layer and Media Access Control (MAC) sub-level of the Data Link layer (Layer 2), takes into account all the possible variables it depends on, which, as will be shown, are different in Mobile WiMAX

and LTE.

Our approach can be considered as belonging to the group of models called *rate-control* models. These models tackle the problem of maximizing data-rate taking into account the combination of system configurations and data-rate provision as a function of the load (rate) distribution [40, 41, 71]. The performance model is then the near-optimum solution to provide the demanded data-rate in the target area within the best system configuration. Although of great interest for 4G networks, however, rate-control models have been less investigated [30, 66, 67, 72] since they have non-linear relations between cell loads [41].

Taking into account these considerations, and aiming at reaching the aforementioned thesis goal, we have considered the following issues:

1. **System Configuration.** We have studied in detail the technological features of LTE and Mobile-WiMAX, considering the different system configuration options. We have modeled the *available resources* for each configuration and *resource formation and allocation* types:
   - As the number of users increases, the operational system overheads to keep track of these users also increase. So, apart from the fixed system overheads and data load consumed by the users, a dynamic overhead calculation is required to know the remaining resources at each moment.
   - The resource allocation mechanisms of the technology under study needs to be mathematically formulated to obtain the available resources under each specific configuration and to know to what extent each technology shares these available resources between its users.
2. **Service Profile.** Those users located in the service area may demand different services, have different signal reception conditions, and consume different amount of system resources. Thus it is necessary for an accurate modeling to elaborate a "Service Profile" definition to model the user's condition and demand.
3. **System Capacity.** An interactive algorithm is required to compare the system *available resources* to the *demanded resources* aiming at determining whether or not the resources can meet the new demands when other extra user tries to enter the system. Figure 1.3 represent a flow chart of the proposed algorithm to compute the capacity in both LTE and Mobile WiMAX systems. It consists of two processes that works in parallel, and estimate, respectively, the amount of real available resources (once those used for PHY+MAC overhead have been removed, using a *system configuration* and a *user distribution profile*) and the amount of resource consumed, using a *service profile*. n is an internal incremental variable that accounts for the number of users. The algorithm ends in providing the capacity as the maximum number of supported users (N) fulfilling the condition that the available resources are greater than the demanded ones.

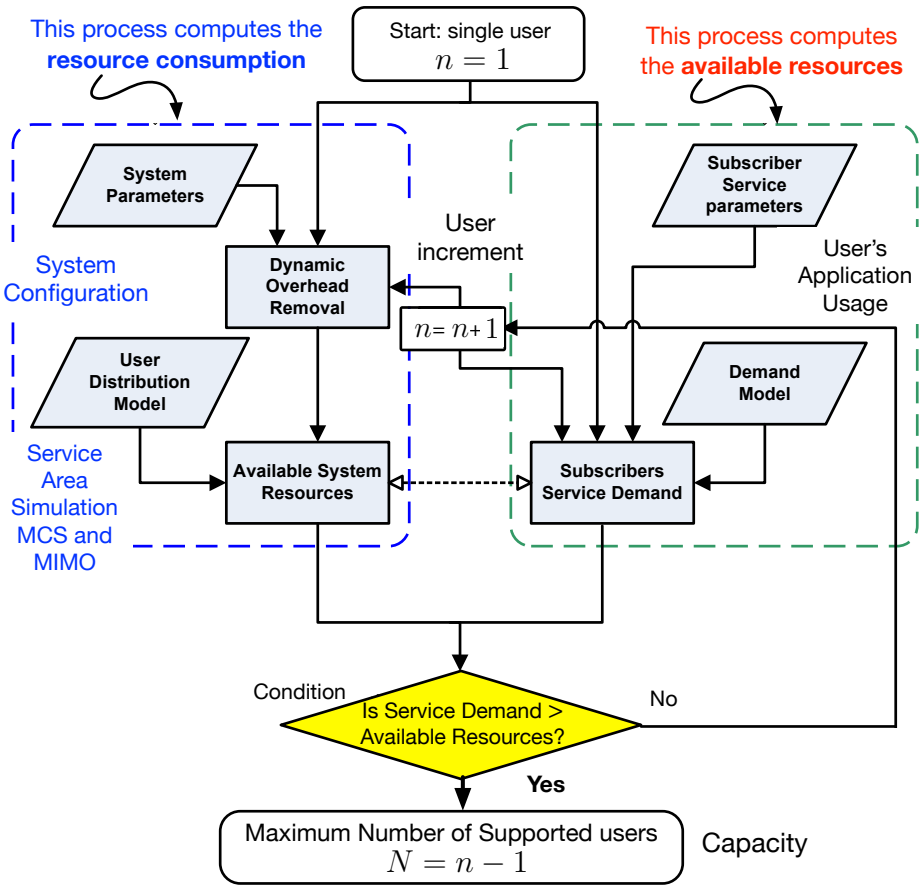


Figure 1.3: Flowchart of the proposed algorithm to compute system capacity. It consists of two processes that works in parallel, and estimate, respectively, the amount of real available resources (once those used for PHY+MAC overhead have been removed, using a *system configuration* and a *service profile*) and the amount of resource consumed. n is an internal variable that accounts for the number of users. The algorithm ends in providing the *capacity* as the maximum number of supported users, N.

1.4 Related work

As motivated, the twofold purpose of this thesis consists in accurately modeling: 1) *useful data throughput* (which requires to compute in parallel the available resources –removing those used for *overhead*– and the demanded resources by several multi-service users), and 2) *capacity* (the maximum number of simultaneous multi-service users that a cell is able to support with a given Quality of Service). Capacity corresponds to the goodput (real throughput once overhead has been removed) of the system at full-load. The obtained capacity can be used as the basis for network design, planning and dimensioning purposes, as well as cost estimation of the new generation cellular networks.

The state of the art related to the thesis purpose is mostly focused on the *capacitating* and *dimensioning* parts. Regarding this, it is worth mentioning that the network development life cycle has several phases. From the technical point of view, the most important phases can be classified into the following general stages: 1) System definition and analysis; 2) System design; and 3) Implementation and testing. Most of the attention of the available literature is mostly drawn on the second phase, the system design. However, in more complex networks, such as mobile networks, the general phase of designing the system can be divided into further subclasses. Among them we can mention: (a) capacitating and dimensioning, (b) operational design, and (c) performance enhancement. Our study in this thesis is mostly focused on the *capacitating* and *dimensioning* parts, and covers their extensive concerns. In a cellular network such as Mobile WiMAX and LTE, by knowing how many users can be supported by each specific hotspot, the number of required nodes to properly cover the service area can be calculated.

Focusing first on Mobile WiMAX capacity studies, the previous, partially related work in the literature are mostly focused on operational and enhancing phases of designing a WiMAX system. As far as we know, none of these studies are concerned with the twofold purpose of this thesis. Therefore, this section reviews the related works that, although focusing on operational issues, play a key role in our capacitating and dimensioning study.

Regarding this, a simple *capacity estimation method* for WiMAX systems has been studied in [73], considering an OFDM-FDD frame structure. The cell capacity results have been presented as a function of the number of users, but considering only two types of traffic services (real-time and FTP). The authors conclude that the capacity is highly correlated to the number of active users in the system, their service types and the frequency reuse pattern. However, the calculations in this paper are OFDM based and the vast details of TDD-OFDMA configurations are *not* discussed.

One key element in capacity estimation consists in properly defining a *service delivery model* for the system under study. The service delivery model requires the knowledge of available system resources, along with the resource consumption, also known as "system load". In Mobile WiMAX networks, the users may have different

sets of traffic with different QoS requirements, be in different receiving conditions, and use different coding and modulation schemes. Consequently, they may consume quite different amounts of the system resources. Accordingly, the overhead required to accommodate the users signaling and control information may vary, resulting in varying amount of available resources. Hence, characterizing a service delivery model turns into a *complex* problem.

The study done in [74] as a part of the Application Work Group of WiMAX Forum, aims to model a consistent input for application-level profiles. The authors provide statistical profiles that include traffic-mix ratio, data-session attempts for applications, and diurnal-application traffic distribution. Regarding this, in our work described next chapters, we use these data to define a simplified application profile for our capacity estimation algorithm, although the implementation of the detailed profile will be straightforward. The aforementioned paper [74] also studies the system performance for VoIP and TCP applications, by presenting the maximum number of users that the system can support for each of those applications. However, they do *not* consider the co-existence of different applications. Furthermore, the presented results correspond to those cases in which the entire frame is utilizing a single MCS level, which means that the link adaption key feature has *not* been considered. Other shortcomings of this study are that the resource overheads are roughly estimated and that the impact of burst construction limitations in resource allocation is *not* considered.

The authors in [75] discuss a conditional optimization problem in order to minimize the required amount of system shared resources, while the QoS requirements of all the scheduled users are satisfied. This problem can be seen as a generic system load model in which either the sum of *time-frequency resources* consumed by all the users (r_{DL}) or the sum of transmission power values consumed by all the users (p_{DL}) is considered as the restricting factor. In other words, the system will be *fully loaded* when *one* of these two factors reaches its *maximum available value*. The *system load* is optimal when resource/power allocation is minimized with respect to the set of modulation and coding schemes ($\mathcal{Q}$), the set of transmission power values ($\mathcal{P}$) and the set of frequency sub-channels ($\mathcal{F}$) that are assigned to all the users in downlink,

$$u_{\mathrm{DL}} = \min(f\{\max(r_{\mathrm{DL}}, p_{\mathrm{DL}})\}), \tag{1.1}$$

where r_{DL} and p_{DL} represent, respectively, the corresponding normalized values (i.e., sum value divided by max value) for time-frequency resources and power. However, as the authors of [75], the solution to this problem is an exhaustive search and is inapplicable in practice.

An *iterative* solution for the aforementioned optimization problem has been discussed in [76] for WiMAX systems, using frequency diversity frame structure (i.e., PUSC). In this reference, the authors present a heuristic algorithms that aims to minimize the sum of allocated resources (r_{DL}) in a frame, while considering that the resource and power constraints *fulfill* the QoS requirements of the scheduled service flows. Note that because of the frequency diversity, the sub-channel con-

straint $\mathcal{F}$ is eliminated. The iteration initiates by applying the highest MCS (i.e., the highest in $\mathcal{Q}$ set) to all the services, which implies the optimal use of the available resources and, accordingly, the highest transmission power. If the restricting factors are exceeded, the algorithm selects one service flow and reduces its MCS level, thus updating the $\mathcal{Q}$ set for the next iteration. The algorithm may iteratively discard one service flow, in those cases in which the explored sets of $\mathcal{Q}$ do not fulfill the constraints. The presented solution imposes a high level of implementation complexities. The real-time simulation of the presented algorithm involves massive data correlations that make it inappropriate for network planning and capacitating studies. The algorithm requires to keep track of the time-varying consumed transmission power of all users, which, in practice, are randomly distributed within the coverage area. Apart from this high complexity, the explored algorithm contains a number of *impractical* assumptions. The number of scheduled services must be known in advance, and discarding the service flows in those iterations in which the constraints are not satisfied means increasing the dropping probability. The resource allocation is performed over the complete downlink subframe, and the fixed and variable resource overheads are *not* taken into account. The resource allocation in downlink subframe is assumed as contentious horizontal strips, which is against the burst construction strategies defined by the standard.

The previous review suggests that there is no study that focuses on calculating the capacity of WiMAX system by proposing the algorithm whose flowchart has been represented in Figure 1.3. It compares the system *available resources* to the *demanded resources* aiming at determining whether or not the resources can meet the new demands when other extra user tries to enter the system. This work, published in [69], will be described in Chapter 2.

On the other hand, when focusing on the related papers in LTE, performance models can be classified into two basic categories: *power-control* based models and *rate-control* models. The first one, formed by *power-control* based models, focuses on power optimization by taking into account both the system configurations and the service area as a function of Signal-to-Interference-plus-Noise Ratio (SINR) distributions [40,77]. Usually the involved powers are variables in optimizing objective functions (Channel Quality, Quality of Service –QoS–, etc.) which, in turn, are functions of SINR. There are several approaches for power optimization involving non-linear objective functions in which an interference matrix contains non-linear power coupling elements [78–81]. The second category of performance models, called *rate-control* models, tackles the problem of maximizing data-rate taking into account the combination of system configurations and data-rate provision as a function of the load (rate) distribution. The model we propose in this paper belongs to this rate-control approach. In contrast to power-control models, the service requirement in rate-control models is not an SINR threshold, but a required data volume to be served during a given time period [40, 41, 71]. The performance model is then the near-optimum solution to provide the demanded data-rate in the target area within the best system configuration. Although of great interest for LTE networks, however, rate-control models have been less investigated [30, 66, 67, 72] since they have

non-linear relations between cell loads [41].

Among the research papers on LTE data-rate performance models, overhead is usually estimated as either a constant value (regardless of operation conditions) [67], or as a percentage of the total available resources [30, 57, 66, 72], or even assuming some overhead mechanisms as negligible [82]. Regarding this, the data-rate model in [82] aims at evaluating LTE performance, based on a clear overview of the basic LTE features, including its different transmission schemes. However, due to its introductory character, the fundamental concept of resource allocation in LTE has *not* been analyzed in detail. As a consequence, the model does *not* consider that overhead depends on the vast number of possible configuration flexibilities.

As mentioned, one of the approaches found in the literature consists in assuming overhead as a constant value [67, 83]. Specifically, the authors in [67] have expanded the peak data-rate model with an overhead count for a single user, although assuming a constant number of overhead controlling symbols per frame. Specifically, the results described in this work correspond to 2 particular cases of 1×1 and 4×4 schemes (n-Tx $\times$ m-Rx) for the lowest and highest bandwidths and for a single user, concluding that the throughput gain is proportional to the bandwidth expansion. The dynamic overhead calculation and multi-user throughput scenario with adaptive multi-antenna transmission techniques per user have *not* been considered.

Another interesting group of data-rate based models are those that assume overhead as a percentage of the total amount of available resources [30, 57, 66, 72, 84]. Specifically, the data-rate performance model explored in [72] aims at estimating LTE peak data-rate by considering that the transmitted data volume is equal to the highest Transport Block Size (the one delivered by the MAC sub-level to the PHY level during 1 TTI) allowed for a single user. RS overhead has been estimated as a percentage of the total amount of RE: 4.8%, 9.5% and 14.3% for 1,2, and 4 transmitting antennas respectively [72]. In a similar approach, the influence of overhead on data rate has also been studied in [66], which focuses on describing a system simulator that includes overhead for signaling and controlling as percentages of the total resources. In particular, in DL, RS overhead have been assumed to be 4.55%, 9.09%, and 12.12% for 1, 2, and 4 receiving antenna, respectively, while control overhead is assumed to be 1-3 symbols per subframe (about 12-21%). A third and more recent work is [30], which explores the feasibility of using LTE for smart grids. It makes use of a model that takes into account the number of REs available for the PDCCH (within a downlink sub-frame) by subtracting those REs reserved for Physical Control Format Indicator Channel (PCFICH), Physical Hybrid-ARQ Indicator Channel (PHICH), and RSs. However, SSs, at frame scale, have *not* been used in this work. Also in the approach of modeling overhead as a percentage of total resources, [84] has recently investigated the performance of signaling overhead caused by keep-alive messages (generated from always-on applications) based on LTE Radio Resource Control (RRC) state transition method for various RRC inactivity timer settings, although without considering the handover related signaling.

A key point to note in the works reviewed so far is that overhead, as we under-

stand it, refers to those REs that are used to carry out signaling, control and synchronization information which involve tasks at PHI level and MAC sub-level. There are however other research papers that, using a number of different viewpoints, tackle particular aspects related to overhead, such as in tracking and paging [85–87], handover [88, 89], non-conventional CQI reporting schemes with reduced overhead [90], core network signaling overhead [91], and those features related to overhead caused by protocols' *header information* in the levels above MAC sublayer [83, 92]. Regarding this miscellanea of heterogeneous papers with different approaches, all the overhead-headers added by protocols in layers above the MAC sublayer will be assumed in our approach as *part of the protocol data unit* (PDU) that the MAC sublayer receives. Examples of protocols (above the MAC sublayer) very commonly used in LTE are: Hypertext Transfer Protocol (HTTP) and Real-time Transport Protocol (RTP) in Layer 5 (Application), or Transmission Control Protocol (TCP) and User Datagram Protocol (UDP) in Layer 4 (Transport), or Internet Protocol (IP) in Layer 3 (Network), or RObust Header Compression (ROHC) protocol in Layer 2. For instance, Voice over IP (VoIP) in LTE uses the ROHC protocol for IP/UDP/RTP header-compression, and a semi-persistent scheduling strategy to assign resources to users (adequate for services characterized by periodic and frequent transmission of small packets such as VoIP) [93–96]. TCP/IP protocols are more used for mobile Internet, for which dynamic scheduling (every 1ms TTI) is more used [92]. Regardless of the protocols used for one service or another (and whose overhead-headers are already included in the PDU that the MAC sublayer receives), the overhead at PHY+MAC level described *must always be applied*.

As a conclusion, the reviewed related papers suggest that overhead plays a key role to accurately compute LTE and Mobile WiMAX performances (basically, data throughput and capacity) in a variety of feasible configurations, and despite of this, they have not received much interest in the technical literature to the point that, to the best of our knowledge, there are *no* models containing all components and dependencies. In view of the lack of an in-depth and comprehensive capacity and data throughput-based performance model including the influence of all PHY + MAC -related overhead contributions, we present the the thesis structure that follows.

1.5 Thesis structure

Based on the arguments in the previous section, the rest of this thesis is organized as follows:

Chapter 2. We have just mentioned that the maximum number of users that a mobile access point can support is a key parameter from both technical and economical viewpoints, especially in a context of competence and financial and economical crisis that does not seem to be over yet clearly. Specifically, in Mobile WiMAX, this number is not easy to calculate because the Mobile Profile of WiMAX technology, based on IEEE 802.16e standard, is able to support

multi-application services along with a wide variety of implementation flexibilities. This rich variety of operation configurations makes Mobile WiMAX more difficult to model than other technologies. Regarding this, we describe in this chapter a simple although flexible, performance data-rate method that we published in [69] to estimate the maximum number of simultaneous multi-service users that each specific Mobile WiMAX access point can support. When we published our research [69] in 2012 there was two main technologies competing for the International Mobile Telecommunications (IMT)-Advanced initiative, WiMAX and LTE, and it was not clear at that moment which of the two would win and lead the Broadband Mobile Internet [56]. The term capacity we have used in [69] must be understood as the upper limit number of users imposed by mandatory overhead used at Physical (PHY) layer and Medium Access Control (MAC) sub-layer. Overhead resources are necessary for controlling, signaling and synchronization tasks at the PHY and MAC layers, and play a key role to accurately compute performance (basically, data goodput) in a variety of possible configurations, despite not having received much interest in the technical literature to the point that, to the best of our knowledge, there is *no* complete model containing all components and dependencies. Specifically, one of the key aspect of our method consists in dynamically computing all PHY+MAC overheads along with formulating the corresponding overhead removal method which allows for computing an improved system *goodput* (or real data throughput after overhead removal) with respect to users distribution and multi-burst construction strategy. To do this properly, the method also contains a service delivery model that investigates the QoS requirements of the services to be supported, along with elaborating an application profile as a consistent input for capacitating and dimensioning studies. Thus the method allows for estimating the minimum resource consumption for the joint-application users at each given time. The core of our method consists of an incremental algorithm which compares the optimal available resources with minimum service demand for each number of users to arrive at maximum system capacity. As will be shown throughout this chapter, by using the proposed algorithm, we have studied different simulation scenarios based on the most used WiMAX implementation parameters in practice. The simulation results prove the significant roll of overhead calculation in performance evaluation studies. Furthermore, these results can be used for network planning and dimensioning purposes, as well as providing reference measures for scheduling performance analysis and detailed simulations.

Chapter 3. Overhead resource elements in Long Term Evolution (LTE) networks are used for some control, signaling and synchronization tasks at both the Physical (PHY) level and Media Access Control (MAC) sub-level. This chapter describes a comprehensive downlink data rate-based performance model that aims at accurately computing *all* the overhead REs (and thus their corresponding influence on data throughput, which, otherwise, would be overestimated), within a unified framework including all the involved variables (bandwidth,

resource allocation type, multiple antenna technology, Hybrid Automatic Repeat Request configuration, and number and reception conditions of users). The model also allows for *dynamically*: 1) optimizing data throughput by selecting, when possible, the configuration that minimizes overhead (not only in single user scenarios, but also in multi-service, multi-user scenarios); and 2) computing the maximum number of users per Transmission Time Interval, the smallest interval in which LTE assigns resources to users. The model explains why *sometimes* increasing the number of antennas in Multiple-input Multiple-output (MIMO) schemes may lead to degrade bit rate rather than increase it. This counterintuitive result is because of the additional overhead resources required by higher order transmission schemes, which reduces the number of resource elements available for user data.

Chapter 4. This chapter discusses the main conclusions obtained. Although, when we began this thesis in 2012 there was two main technologies competing for the International Mobile Telecommunications (IMT)-Advanced initiative –WiMAX and LTE–, it was not clear at that moment which of the two would win and lead the Broadband Mobile Internet. However, at present, the situation is clearer: LTE seems to be the *undisputed* winner for *broadband cellular* implementations, while WiMAX has now been confined to very specific market niches such as aviation industry, smart grids, point-to-point long distance and high rate radio transmission, instead of point-to-mutiple-point cellular environments. In any case, what LTE and Mobile-WiMAX do have in common is that they are both very complex solutions which, apart from technological similarities (such as use of OFDM technology), have many structural and implementation differences. Additionally, the standards that define these two broadband mobile solutions, both contain numerous flexibilities. The configuration choices are left open for the vendors to find their optimal solution according to their service area. The performance of these networks is therefore highly dependent on solving this complex optimized design problem i.e. system vs. load. The network planning and optimize design in new generation mobile networks is an active research line, specially in LTE, the clear winner. High-level and, at the same time, accurate performance modeling are of high value in planning cellular networks, as full scale dynamic simulations are not affordable in time and complication for large planning scenarios.

Chapter 2

Exploring the Maximum Number of Simultaneous Multi-service Users in Mobile WiMAX

We have mentioned in Chapter 1 that the maximum number of simultaneous, multi-service users that a mobile access point can support is a key parameter from both technical and economical viewpoints, especially in a context of competence and financial and economical crisis that does not seem to be over yet clearly. Specifically, in Mobile WiMAX, this number is difficult to calculate accurately because the Mobile Profile of WiMAX technology, based on IEEE 802.16e standard, is able to support multi-application services along with a wide variety of *implementation flexibilities*. This rich variety of operation configurations makes Mobile WiMAX more difficult to model than other technologies [56, 69].

Regarding this, we describe in this chapter a simple, although flexible performance data-rate method that we have published in [69] to estimate the maximum number of simultaneous multi-service users that each specific Mobile WiMAX access point can support. When we published our research [69] in 2012 there was two main technologies competing for the International Mobile Telecommunications (IMT)-Advanced initiative [97], WiMAX and LTE, and it was not clear at that moment which of the two would win and lead the Broadband Mobile Internet [56, 98]. The term "capacity" we have used in [69] must be understood as the upper limit or maximum number of users imposed by mandatory overhead used at Physical (PHY) layer and Medium Access Control (MAC) sub-layer. In both WiMAX and LTE systems, overhead resources are necessary for controlling, signaling and synchronization tasks at the PHY and MAC layers, and play a key role to accurately compute performance in a variety of possible configurations, despite not having received much interest in the technical literature to the point that, to the best of our knowledge, there is *no* complete model containing all components and dependencies. Specifically, one of the key aspect of our method consists in dynamically computing all PHY+MAC overheads along with formulating the corresponding overhead removal method which allows for computing an improved system *goodput* (or real data throughput after

overhead removal) with respect to users distribution and multi-burst[1] construction strategy. To do this properly, the method also contains a *service delivery model* that takes into account the Quality of Service (QoS) requirements of those services to be supported, along with an *application profile* as a consistent input for capacitating and dimensioning studies. Thus the method allows for estimating the minimum resource consumption for the joint-application users at each given time. The core of our method consists of an incremental algorithm which compares the optimal available resources with minimum service demand for each number of users to arrive at maximum system capacity. As will be shown throughout this chapter, by using the proposed algorithm, we have studied different simulation scenarios based on the most used WiMAX implementation parameters in practice. The simulation results prove the significant roll of overhead calculation in performance evaluation studies. Furthermore, these results can be used for network planning and dimensioning purposes, as well as providing reference measures for scheduling performance analysis and detailed simulations.

2.1 Introduction

Mobile WiMAX, based on IEEE 802.16e standard [68], is one of the two main technologies that started to compete for the IMT-Advanced initiative. In fact, Mobile WiMAX is part of a broader set of standards grouped under the name IEEE 802.16. This is not an unique standard but also a family of standards which, focused on the radio interface (air interface), aims at providing a wide variety of broadband wireless services along with product certification. IEEE 802.16 is usually named WiMAX. In particular, within this framework, the standard known IEEE 802.16-2004 [99] has been developed to add Non-Line of Sight applications for being used in fixed and nomadic networks, while IEEE 802.16e-2005 amendment [100] has been designed for supporting mobility requirements for a variety of applications.

The WiMAX Forum states a "mobile system profile" [101] for Mobile WiMAX technology interoperability, which is based on a number of mandatory and optional standards, focused on key features in the PHY (Physical) layer and MAC (Medium Access Control) sublayer, in a Point to Multiple Point (PMP) operating mode. However, this profile contains a wide range of *implementation flexibilities* so that some decisions such as the choice of scheduling methods or the resource allocation techniques are left up to the vendors. Therefore, there have been many studies on its technical improvements and service provision capabilities [73, 74, 102–106].

Nevertheless, WiMAX capacity analysis is still remained as a complex problem. Just in this respect, and in the effort of developing a capacity estimation methodology for novel generation wireless networks, there are many aspects, ranging from air-link to applications, which must be integrated. For WiMAX systems these figures include, but are not limited to:

[1] As presented in Chapter 1, a "burst" or data region is a two-dimensional (time, frequency) allocation of slots (the minimum unit to be allocated to a user) in the OFDMA frame.

1. System characteristics [54–61, 63, 68, 100, 101, 107–111]
2. Service provision strategies [104, 112–117],
3. Coverage model [104]
4. Load distribution [73–75, 102, 104–106, 112–116]
5. Quality of Service analysis [62, 118, 119]
6. Forecasting mobile broadband traffic [120]
7. Energy saving considerations [121–124]
8. Scheduling issues [122, 125–134]
9. Comparative studies between Mobile WiMAX and LTE [57, 107, 135–139]
10. Integration of LTE and mobile WiMAX networks [140–142]

Nota that any of the aforementioned topics contains considerable implementation complexities so that the evaluation of the complete system performance requires massive simulations [74, 103, 143]. For further details about key issues related to system-level modeling of Mobile WiMAX networks the interested reader is referred to [104].

Regarding these considerations, the purpose of this chapter is to summarize the extensive multiple elements involved in the capacity estimation problem and solve this problem for Mobile WiMAX systems. In this chapter, the term capacity refers to the maximum number of simultaneous multi-service users that each specific Mobile WiMAX access point can jointly support (cell goodput[2]). We have developed a dynamic PHY+MAC overhead computation method which is able to achieve an improved system goodput with respect to users distribution and multi-burst construction strategy. Our solution focuses on the practical capacity provision capabilities of a WiMAX hotspot by considering the operational details. Specifically, we propose a dynamic and simple algorithm that, at each given condition and for each number of users, calculates the system available resources and compares it with users resource consumption. As long as there is enough resource to support additional users with their QoS requirements, the algorithm increments number of supported users and, at last, finds the largest number of supported users of the hotspot that is characterized with input parameters. This way, the network planner has the key data that is required to solve the optimization problem of covering a specific service area. These information are: (a) which type of access point to be used (determined by the specific input parameters of each sector) and (b) how many hotspot of that type to be implemented (determined by capacity and number of users to be supported).

[2]In a general context, *goodput* refers to the application-level data throughput. The amount of data considered excludes protocol overhead bits as well as retransmitted data packets. Because of that, the "goodput" is lower than the throughput.

Furthermore, by knowledge of the two key features above, the operators can have a good approximation for investment estimation of network deployment.

The rest of this chapter is organized as follows. Section 2.2 focuses on selecting and explaining those key figures of Mobile WiMAX that will assist us in clearly describing our model. With such concepts, Section 2.3 presents an overview of the previous works related to capacity estimation for WiMAX systems. For the sake of clarity, Section 2.4 introduces a previous overview of our capacity estimation algorithm. This will assist us in acquiring a global picture of the concepts involved. To proceed further with the deepest details involved, Section 2.5 focuses on describing PHY+MAC overhead components, and in modeling their influence on the system's available resources, by formulating a PHY+MAC overhead removal methodology, with a look over resource allocation strategies. In Section 2.6 we present a system delivery model based on data-rate requirements of different services to be supported and elaboration of a statistical application profile. Computer simulations and numerical results are presented in Section 2.7 followed by a conclusion at last section. For the sake of clarity, Table 2.1 lists the acronyms used in this chapter.

Table 2.1: List of acronyms used in this chapter.

AMC	Adaptive Modulation and Coding
ARQ	Automatic Repeat Request
BE	Best Effort
BLER	Block Error Rate
BS	Base Station
CBR	Constant Bit Rate
CC	Convolutional Coding
CID	Channel IDentifier
CR	Contention Ratio
CRC	Cyclic Redundancy Check
CTC	Convolutional Turbo Coding
DCD/UCD	Downlink/Uplink Channel Descriptor
DL	Downlink
ErtPS	Extended Real-time Polling Service
IE	Information Element
MCS	Modulation and Coding Scheme
MS	Mobile Station
nrtPS	Non Real-time Polling Service
PUSC	Partially Used Sub-Carrier
QoS	Quality of Service
RRM	Radio Resource Management
RTG/TTG	Receive/Transmit Transition Gap
rtPS	Real-time Polling Service
SDU/PDU	Service/Protocol Data Unit
SNIR	Signal to Noise plus Interference Ratio
SOFDMA	Scalable Orthogonal Frequency Division Multiple Access
TDD/FDD	Time/Frequency Division Duplex
UGS	Unsolicited Grant Service
UL	Uplink
VBR	Variable Bit Rate

2.2 Technical Background

In this section we present a review on the main aspects of the Mobile WiMAX standard that play a role in our capacity estimation methodology. First, two subsections review the PHY features of the standard, while the third subsection describes the data mapping procedure from higher layers into MAC and PHY levels. The last

two subsections focus on the features of the WiMAX technology that are essential for Radio Resource Management (RRM) strategies.

2.2.1 An overview of Mobile WiMAX frame

The standard [68] defines both Time Division Duplex (TDD) and Frequency Division Duplex (FDD) modes for the PHY layer, which uses Orthogonal Frequency Division Multiple Access (OFDMA) technique. In this chapter we have considered TDD-OFDMA since, when we published this work [69], the first releases were expected to use only TDD.

Figure 2.1 will assist us in introducing the main concepts involved. It basically represents a simplified structure of a TDD-OFDMA *frame*, which is a two-dimensional grid in time (t) and frequency (f).

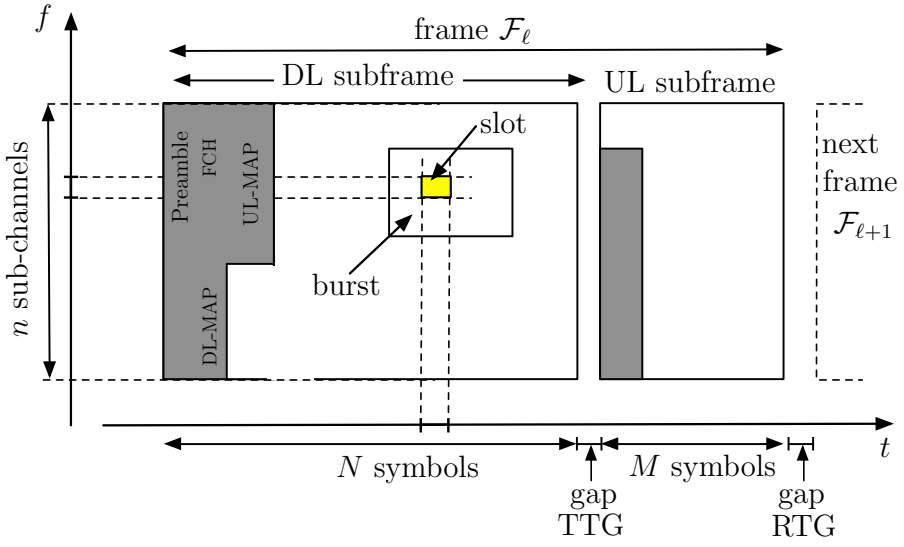


Figure 2.1: Simplified representation of the OFDMA-TDD Frame structure in Mobile WiMAX. TTG and RTG stand for "Transmit Transition Gap" and "Receive Transition Gap", respectively.

Specifically:

- In time dimension, the frame basically consists of a downlink (DL) subframe and uplink (UL) subframe [56]. Each of them transports a number of OFDM symbols, whose details we postpone to Subsection 2.2.2.

- In frequency dimension, the frame consists of a number of orthogonal subcarriers (SCs) grouped into "sub-channels", whose details will be shown in Subsection 2.2.3.

This 2-dimensional frame can be considered as formed by "elemental units" in time and frequency called "slots". A "slot" (yellow rectangle in Figure 2.1) is the *minimum* possible *data allocation* unit (MAU) in the 802.16 standard. Most of the frame slots (white areas in Figure 2.1) are used to transmit users' data. Each of these user's *data region* is called "burst" and is a two-dimensional allocation of slots, that is, a group of contiguous sub-channels transporting a set of contiguous OFDMA symbols [70]. However, part of the frame (gray area) must be reserved for the proper system operation (synchronization, signaling, and control). Although it will be explained in detail later on, we motivate here the need form some of these elements. For instance, a "preamble" is used for time *synchronization*, while the "downlink map" (DL-MAP) and "uplink map" (UL-MAP) are used to communicate both the burst-start time and burst-end time, along with the modulation schemes and forward error control (FEC) technique for each mobile station (MS). The other remaining element in Figure 2.1, the "Frame Control Header" (FCH), is used to specify the lengths and subcarriers to be used in these MAPs.

With this introduction in mind, Subsections 2.2.2 and 2.2.3 focuses on describing time and frequency domains details, respectively.

2.2.2 The Time Domain in the Frame Structure

In time dimension, frame $\mathcal{F}_\ell$ has a length of 5 ms. and is divided into a DL and UL subframes. As represented in Figure 2.1, the DL subframe and the UL subframe within frame $\mathcal{F}_\ell$ are separated by a gap time whose name is "Transmit Transition Gap" (TTG). The gap between frame $\mathcal{F}_\ell$ and the next one, frame $\mathcal{F}_{\ell+1}$ is called "Receive Transition Gap" (RTG). A given frame $\mathcal{F}_\ell$ contains a number N of OFDMA symbols in the DL subframe, and M OFDMA symbols in the UL subframe, whose important details for the purpose of this chapter will be shown later on. The complete frame consists of 47 OFDM symbols, for the aforesaid bandwidths (transition gaps excluded).

A key point to note is that the duration of both the DL subframe (N OFDMA symbols) and that of the UL subframe (M symbols) is *adjustable* (to support asymmetric traffic). N and M can take different allowed values constrained to the restriction $N + M = 47$ symbols. The way the $N + M$ symbols are distributed is represented by a "DL:UL symbol ratio". For instance, 35 symbols in DL and 12 ("35/12" for UL for the ratio of 3:1). This strategy leads to the following allowed DL/UP ratios: $35/12, 34/13, ..., N/M, \cdots, 26/21$, where $N + M = 47$ symbols.

Aiming at mitigating the Inter Symbol Interference (ISI), a Cyclic Prefix (CP) has been introduced in the standard [68, 143]: it copies a portion of the ending of the useful symbol to the beginning of the OFDMA signal wave. If T_b represents the

useful symbol duration, and T_G the corresponding "Guard Time", then the OFDMA symbol duration will be $T_S = T_b + T_G$. The guard time $T_G = T_d/G$ increased with an order of $1/G$, where G can have a value of 4, 8, 16 or 32 [68, 143]. As long as the CP duration is larger than the channel delay spread, the ISI is completely eliminated [68, 143].

Once we know the most important details of the frame in the temporal dimension, the goal now is to explore the key details in the frequency domain that will assist is explaining our capacity model.

2.2.3 Frequency Domain and Subchannelization in Frame Structure

As mentioned in Subsection 2.2.1 when introducing the OFDMA-TDD frame, in the frequency domain, the total number of OFDMA sub-carriers are grouped into a number of "sub-channels". Besides this, another key element in Mobile WiMAX is the fact that its OFDMA scheme is *scalable*. Although this specific word has not been used in the standard, OFDMA PHY is said to have "Scalable" OFDMA (SOFDMA). The scalability is the change of the FFT size, and thus the number of subcarriers. The mandatory supported FFT sizes for mobile WiMAX profiles are 1024 and 512.

The total number of available subcarriers per symbol is determined by the FFT size, which in turn depends on the system bandwidth BW. The scalability in OFDMA-PHY in Mobile WiMAX implies that, as the system bandwidth increases, the FFT size increases correspondingly so that the subcarrier spacing remains fixed on 10.97 kHz. This value is chosen to maintain the subcarriers orthogonally, and results in a fixed symbol duration ($T_b = 91.4$ μs, $G = 1/8$).

Among all the available subcarriers, not all of the subcarriers are assigned to carry data. There are a number of guard subcarriers and a DC subcarrier that do not transmit any data (Null subcarriers) and a number of pilot subcarriers that are distributed within the symbols based on the applied permutation mode. In this respect, subcarriers are allocated using the so-called "OFDMA permutation modes", which are basically as follow:

- Diversity (or distributed) permutations. Subcarriers are distributed pseudo-randomly. This type includes, among others, the PUSC (Partial Usage of the SubChannels) permutation, which is mandatory in both DL and UL. The main advantages of distributed permutations are frequency diversity and intercell interference averaging.
- Contiguous (or adjacent) permutations. These consider a group of adjacent subcarriers. This family includes the AMC (Adaptive Modulation and Coding) mode. This type of permutation leaves the door open for the choice of the best-conditions part of the bandwidth. Channel estimation is easier as the subcarriers are adjacent. AMC is mandatory for both DL and UL.

Specifically, the mandatory PUSC permutation mode divides pseudo-randomly the OFDMA subcarriers into a number of sub-channels with 48 SCs. Thus, PUSC is a distributed permutation mode that results in the desired frequency diversity for mobile communications. Table 2.2 illustrates the PUSC mode characterization for 5 and 10 MHz channel widths.

Table 2.2: Summary of SOFDMA PHY parameters in PUSC mode as a function of the bandwidth BW. SC means subcarrier.

BW	**FFT Size**	**No. Data SCs**		**No. Pilot+Null SCs**		**No. Sub-Channels**	
		DL	**UL**	**DL**	**UL**	**DL**	**UL**
5 MHz	512	360	272	60+92	136+104	15	17
10 MHz	1024	720	560	120+184	280+184	30	35

To complete this subsection focuses on frequency-related aspects, it is worth mentioning that different Frequency Reuse Patterns (FRPs) can be implemented over frequency resources of a WiMAX frame. A (1,1,3) pattern divides the available BW of a cell over 3 sectors (1/3 BW per sector), while in the (1,3,3) pattern, one frequency band is available for each sector (BW per sector). However, some improvements to frequency reuse schemes are available [103].

2.2.4 Data Mapping

Each IP Packet is associated to an MAC SDU (Service Data Unit). In the presence of ARQ (Automatic Repeat reQuest) mechanism, the transmitter may divide each SDU into fragments, whose lengths are specified by using the ARQ-BLOCK-SIZE parameter [70, 102, 144, 145]. The value of this parameter (in number of bytes) is set by a negotiation process during the connection creation dialog [144, 146]. The sets of the corresponding ARQ blocks, which have been selected for transmission, have then to be encapsulated forming a MAC PDU (Protocol Data Unit), whose main characteristics for the purpose of this chapter are as follows. First, the formation of this element MAC-level PDU requires to add a 6-byte generic *header* at the beginning of the received set of blocks, and also a 32-bit CRC block at the end. Additionally, a fragmentation sub-header is inserted in the PDU, if all the ARQ blocks are contiguous and related to the same SDU. If this is not the case, a packetization sub-header appears between two adjacent ARQ blocks that belong to different SDUs and are encapsulated in a single PDU. In other words, the existence of fragmentation/packetization sub-header in a PDU is a *mutually exclusive* operation. At physical level, each MAC-PDU is partitioned into groups of bytes, which are modulated and coded according to the profile of the corresponding burst allocated for transmission.

2.2.5 Resource Allocation

The minimum possible data allocation unit (MAU), assigned to a connection by the scheduler, is, as mentioned before in Subsection 2.2.1, a slot, which is a two-dimensional entity in time and frequency. The number of available slots in either the DL or UL subframes depends on: (a) the number of OFDMA symbols in each subframe, specified by the DL:UL symbol ratio; and (b) the applied subcarrier Permutation Mode. Using PUSC permutation, the 48 data subcarriers of a DL sub-channel are equally divided over 2 OFDMA symbols (24+24), while, in UL, the subcarriers are spread over 3 symbols following a (12+24+12) scheme [70,103]. Each slot may carry different amounts of data-bits depending on the applied modulation and FEC (Forward Error Correction) channel coding level, resulting in different data-block-size per sub-channel. IEEE 802.16e standard supports QPSK, 16-QAM and 64-QAM modulation with various rates of convolutional codes (CC) or turbo codes (CTC), while the support for 64-QAM is optional in uplink. Table 2.3 lists the block size (bits) related to each Modulation and Coding Scheme (MCS) in PUSC mode.

Table 2.3: Block-size (bits) as a function of the Modulation and Coding Scheme (MCS) in PUSC (Partially Used SubCarrier) permutation mode.

Tab	**Modulation**	**CTC**	**Block Size (bit)**	**Min SNR**	**Weight**
1	QPSK	1/2	48	2.9	0.20
2	QPSK	3/4	72	6.3	0.17
3	16-QAM	1/2	96	8.6	0.20
4	16-QAM	3/4	144	12.7	0.13
5	64-QAM	1/2	144	13.8	0.15
6	64-QAM	2/3	192	16.9	0.10
7	64-QAM	3/4	216	18	0.05

Figure 2.2 will assist us in explain better some important topics in resource allocation. As represented, a contagious series of slots can be assigned to a given user, called “user data-region”. A data region, commonly depicted as a rectangle, is a two dimensional allocation of a group of contiguous sub-channels (i.e.: with contiguous logical numbers) in a group of contiguous OFDMA symbols. Each data-burst may contain a number of data-regions, and each subframe may contain several burst of varying size (with maximum number of 16 bursts in DL [101]). When a specific DL-burst is meant for a given MS (Mobile Station), the burst properties along with the unique Connection Identifier (CID), are some of the pieces of that compose the corresponding “Information Element” (IE) within the DL-MAP message. Note that each MS may simultaneously request different applications, hence having different

connections and different CIDs. All users track the DL-MAP, so that each user receives the downlink data which is addressed to its specific CID. The DL-MAP-IE informs the user about the rectangular place of the DL resource allocation within the DL subframe, as well as the necessary information for decoding the message. Similarly, the resource allocation in UL is specified with UL-MAP-IEs, although the UL data burst is allocated as a horizontal strip per subscriber. The transmission starts at a particular slot, and continues during the entire UL subframe. The horizontal allocation is used to minimize the number of subcarriers for each MS and hence maximizes the power per subcarrier.

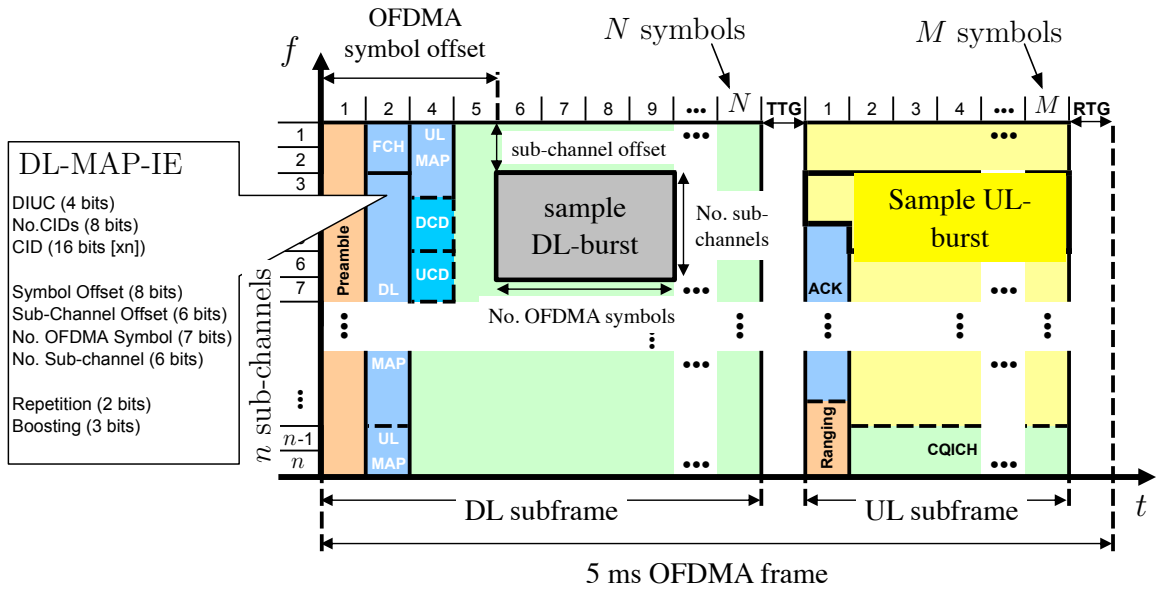


Figure 2.2: OFDMA-TDD Frame structure indicating control messages and resource allocation procedure. See the main text for further details.

We completed the description of Figure 2.2 with the the Downlink Channel Descriptor (DCD) and Uplink Channel Descriptor (UCD) messages, which are sent periodically whenever the link conditions change and new channel profile must be introduced to MSs. The concepts of preamble (for time synchronization), DL-MAP and FCH have already been explained in Subsection 2.2.1.

2.2.6 Application Classes and Quality of Service

The overall applications to be supported with Mobile WiMAX are classified into 5 major categories by WiMAX Forum [147]. These classes are:

1. Interactive Gaming
2. VoIP
3. Streaming
4. Web Browsing
5. File Transfer.

The major metrics for this classification are the data-rate, the latency and the jitter requirements of each application. For each application (i.e., data service), these parameters must be known by the system aiming at guaranteeing a quality user experience. In other words, based on the QoS (Quality of Service) requirements for the different existing data services, the system decides on their transmission ordering and scheduling on the air interface. In WiMAX, this task is done by assigning "Service Flows" to each connection between the BS and the MS at MAC level. In fact, a service flow is a unidirectional flow of packets that is provided with a particular set of QoS parameters. These *service flows*, listed in decreasing serving priority order [147], are:

1. Unsolicited Grant Service (UGS)
2. Real-time Polling Service (rtPS)
3. Extended Real-time Polling Service (ErtPS)
4. Non Real-time Polling Service (nrtPS)
5. Best Effort (BE).

The properties of these services flows will be further studied in Section 2.6.2. Prior to this, what we can say in this chapter point is that a typical service flow management strategy for WiMAX systems is composed of the following steps [117]:

1. An admission control mechanism determines whether or not a new request for a connection can be satisfied according to the remaining free resources
2. A buffer manager controls the buffer size
3. A scheduling mechanism determines the sending sequence of the packets of the admitted connections. The scheduler decides which ones to send and when to send.

Thus, there may be one separate queue for each service flow and each user. The scheduler matches the channel rates (frequency resources) with respect to users channel quality and assigns the transmission slots within the frame (time resources) according to the priority of each service flow. A scheduler design for multiple traffic classes in OFDMA networks has been proposed in [112]. This scheduler has been found to be well suited for WiMAX systems. In the research work [112], the traditional Modified Largest Weighted Delay First (MLWDF) scheduling scheme is extended for multiple traffic classes, where the used scheduling metrics are: Head of line packets waiting time for restricted delay traffic and queue length information for BE traffic. In another study [105], the scheduling has been viewed as a cross-layer scheme for AMC permutation mode. The resources are allocated with respect to the MAC and PHY joint responses for different QoS requirements. Since the scheduling algorithms for WiMAX are not standardized, a considerable number of other studies may be found in the literature. For detailed information, the reader is referred to [106] that summarizes various scheduling strategies and provides performance evaluation comparisons on different methods. For the sake of clarity, in this chapter, we do not simulate a complex smart scheduler, and instead, we have replaced its complexity with a simple *statistical traffic model*, introduced in Section 2.6.2, called "application profile" in Subsection 2.6.2

With this information in mind, we now have enough knowledge to understand what other scientists have done in this research area. This is just the purpose of the following section.

2.3 Related Work

The network development life cycle has several phases. From the technical point of view, the most important phases can be classified into the following general stages:

- System definition and analysis
- System Design
- Implementation and Testing

The attention of the available literature is mostly drawn on the second phase, the system design. However, in more complex networks, such as mobile and wireless networks, the general phase of designing the system can be divided into further subclasses. Among them we can mention: (a) capacitating and dimensioning, (b) operational design, and (c) performance enhancement. Our study in this chapter is mostly focused on the capacitating and dimensioning parts, and covers its extensive concerns. In a wireless network such as Mobile WiMAX, by knowing how many users can be supported by each specific hotspot, the number of required hotspots to properly cover the service area can be calculated.

The previous studies in the literature are mostly focused on operational and enhancing phases of designing a WiMAX system. As far as we know, none of these studies are concerned with the objective of this paper. Therefore, this section reviews the previous works that, although focusing on operational issues, play a role in our capacitating and dimensioning study.

A simple *capacity estimation method* for WiMAX systems has been studied in [73], considering an OFDM-FDD frame structure. The cell capacity results have been presented as a function of the number of users, but considering only two types of traffic services (real-time and FTP). The authors conclude that the capacity is highly correlated to the number of active users in the system, their service types and the frequency reuse pattern. However, the calculations in this paper are OFDM based and the vast details of TDD-OFDMA configurations are *not* discussed.

One key element in capacity estimation consists in properly defining a *service delivery model* for the system under study. The service delivery model requires the knowledge of available system resources, along with the resource consumption, also known as "system load". In Mobile WiMAX networks, the users may have different sets of traffic with different QoS requirements, be in different receiving conditions, and use different coding and modulation schemes. Consequently, they may consume quite different amounts of the system resources. Accordingly, the overhead required to accommodate the users signaling and control information may vary, resulting in varying amount of available resources. Hence, characterizing a service delivery model turns into a *complex* problem.

The study done in [74] as a part of the Application Work Group of WiMAX Forum, aims to model a consistent input for application-level profiles. The authors provide statistical profiles that include traffic-mix ratio, data-session attempts for applications, and diurnal-application traffic distribution. Regarding this, in our work described in this chapter, we use these data to define a simplified application profile for our capacity estimation algorithm, although the implementation of the detailed profile will be straightforward. The aforementioned paper [74] also studies the system performance for VoIP and TCP applications, by presenting the maximum number of users that the system can support for each of those applications. However, they do *not* consider the co-existence of different applications. Furthermore, the presented results correspond to those cases in which the entire frame is utilizing a single MCS level, which means that the link adaption key feature has *not* been considered. Other shortcomings of this study are that the resource overheads are roughly estimated and that the impact of burst construction limitations in resource allocation is *not* considered.

The authors in [75] discuss a conditional optimization problem in order to minimize the required amount of system shared resources, while the QoS requirements of all the scheduled users are satisfied. This problem can be seen as a generic system load model in which either the sum of *time-frequency resources* consumed by all the users (s_{D_L}) or the sum of transmission power values consumed by all the users (ρ_{D_L}) is considered as the restricting factor. In other words, the system will be *fully loaded*

when *one* of these two factors reaches its *maximum available value*. The *system load* is optimal when resource/power allocation is minimized with respect to the set of modulation and coding schemes ($\mathcal{Q}$), the set of transmission power values ($\mathcal{P}$) and the set of frequency sub-channels ($\mathcal{F}$) that are assigned to all the users in downlink subframe:

$$u_{D_L} = \min(f\{\max(s_{D_L}, \rho_{D_L})\}), \tag{2.1}$$

where s_{D_L} and ρ_{D_L} represent, respectively, the corresponding normalized values (i.e., sum value divided by max value) for time-frequency resources and power.

However, as the authors of [75], the solution to this problem is an exhaustive search and is inapplicable in practice.

An iterative solution for the aforementioned optimization problem has been discussed in [76] for WiMAX systems, using frequency diversity frame structure (i.e., PUSC). In this reference, the authors present a heuristic algorithms that aims to minimize the sum of allocated resources (s_{D_L}) in a frame, while considering that the resource and power constraints fulfill the QoS requirements of the scheduled service flows. Note that because of the frequency diversity, the sub-channel constraint $\mathcal{F}$ is eliminated. The iteration initiates by applying the highest MCS (i.e., the highest in $\mathcal{Q}$ set) to all the services, which implies the optimal use of the available resources and, accordingly, the highest transmission power. If the restricting factors are exceeded, the algorithm selects one service flow and reduces its MCS level, thus updating the $\mathcal{Q}$ set for the next iteration. The algorithm may iteratively discard one service flow, in those cases in which the explored sets of $\mathcal{Q}$ do not fulfill the constraints. The presented solution imposes a high level of implementation complexities. The real-time simulation of the presented algorithm involves massive data correlations that make it inappropriate for network planning and capacitating studies. The algorithm requires to keep track of the time-varying consumed transmission power of all users, which, in practice, are randomly distributed within the coverage area. Apart from this high complexity, the explored algorithm contains a number of *impractical* assumptions. The number of scheduled services must be known in advance, and discarding the service flows in those iterations in which the constraints are not satisfied means increasing the dropping probability. The resource allocation is performed over the complete downlink subframe, and the fixed and variable resource overheads are *not* taken into account. The resource allocation in downlink subframe is assumed as contentious horizontal strips, which is against the burst construction strategies defined by the standard.

2.4 Overview of the Proposed Capacity Estimation Algorithm

As mentioned in the motivation of this chapter, the term capacity refers to the *maximum number of simultaneous multi-service users* that can be supported by each

specific BS, what in turn determines the corresponding *cell goodput at full-load*.

Aiming at computing such capacity, we have proposed an *incremental* solution based on the relation stated by Expression (2.1). In our approach, the problem will be divided into two major processes that will be solved in parallel in a variety of different, feasible service provision cases. One of these two processes computes the optimal available resources, while the other one estimates the minimum resource consumption. Note that, in a real case, the absolute solution to the optimization problem stated by Expression (2.1) is not achievable in a *reasonable* computational time. Therefore, with a detailed view of the technical features and with a focus on planning and dimensioning purposes, our proposed algorithm help put into practice the solution to the capacity estimation problem of WiMAX system.

Put it simple, as the number of active users increases, proportional to the requested application, the resource consumption increases and the available resource decreases. Therefore, serving *additional* users requires a dynamic algorithm to carried out an instantaneous comparison of the cell's serving conditions. As illustrated in Figure 2.3, the first block (in the dashed rectangle on left hand side) calculates the *available cell resources* at each given time. The accurate calculation of useful resource requires the knowledge of number of scheduled active users per frame (N_{ActUsers}) as well as their resource allocation pattern. In each step, the number of useful slots (available resources minus those used for overhead, which cannot be used for user data) in the DL-subframe is calculated using the dynamic PHY+MAC overhead removal methodology proposed in Section 2.5. The optimal allocation of the useful slots is derived from modulation and coding distribution at the coverage area. This will be studied in detail in Section 2.5.2. We assume that the mobile terminals are uniformly distributed inside the coverage area and that the BS has enough transmission power to serve all of the distributed terminals. The offered data-rate of the target cell is formulated by studying the burst construction strategies. For each number of users, the *unknown* parameters to the first process are calculated in the parallel process in the second block. Furthermore, this block (right hand side dashed rectangle) is responsible for the calculation of the aggregated minimum required data-rate for multi-service users. As mentioned before, we have elaborated a statistical methodology to be replaced with the complexities of implementation of a fair scheduling procedure. Section 2.6 introduces the considered application profiles and formulates the aggregated data-rate demand as a function of the number of users and based on different application QoS requirements.

As can be seen in Figure 2.3, each time the number of subscribers (N) increments by one, the actual system resource and subscribers minimum service demand are recalculated and compared. The increment continues as long as the remaining resources can satisfy the additional demand imposed by the new user. When the demand exceeds the available system resource, the algorithm introduces the largest N as the measure for system capacity.

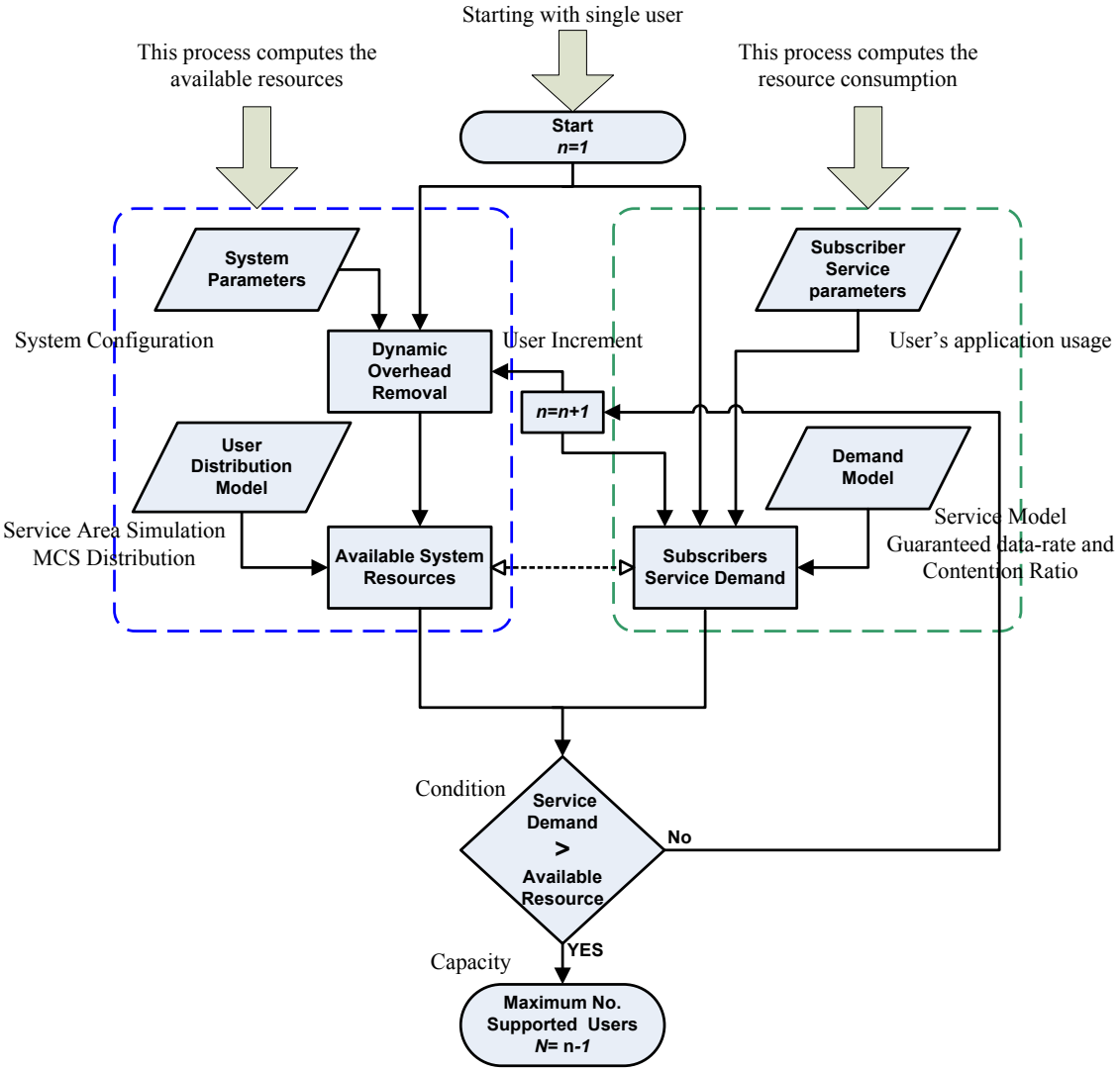


Figure 2.3: Proposed Capacity Estimation Algorithm. It is formed by two processes that works in parallel, and estimates, respectively, the amount of real available resources (once those used for PHY+MAC overhead have been removed) and the amount of resource consumed. n is an internal variable that accounts for the number of users. The algorithm ends in providing the maximum number of supported active users, N.

2.5 Dynamic System Resource Calculation

As we have described before, the system available resource is defined as the time-frequency resources (slots) that are available for data transmission in a Mobile WiMAX frame. As was explained in Section 2.2.3, the resource characteristics of a Mobile WiMAX system depend on the BW and applied Permutation Mode. However, as was explained in Section 2.2.5, the allocation of these resources, and thus the data-rate that they provide, depend on the applied MCS distribution.

Regarding this, we have divided this section into two parts. The first one focuses on computing the PHY+MAC overhead in the DL subframe, in which the remaining resources can be assigned for pure user data transmission. Since overhead resources cannot be used for data transmission, the accurate real data rate must be computed by removing those resources that have to be used for overhead. The second subsection focused on modeling the burst construction in DL-subframe and the real offered data-rate provision by including a modulation and coding distribution assumption.

2.5.1 A model to estimate PHY+MAC overhead the number os useful data symbols (N_{sym})

The synchronization, signaling and control messages and the other transmission overheads, described in Section 2.2, are necessary to maintain the proper functionality of the WiMAX system. Overhead components do not carry user data. In this section we review the most significant overhead components in PHY layer along with presenting the detailed overhead removal methodology in a dynamic manner. In order to further improve the obtained PHY throughput, we propose novel methods to extract the MAC layer overheads and estimate other wasted resources in dynamic allocation.

Returning again to Figure 2.2, a DL subframe starts with one-symbol preamble, thus introducing a fixed overhead in DL. A 24-bit Frame Control Header (FCH) follows the preamble, and is sent with QPSK$\frac{1}{2}$ and 4 repetitions, leading to occupy the 2 first sub-channels in the first data-symbol. This overhead process uses $24 \times 4 = 96$ bits. The next symbol in the frame contains a "DL-MAP message", which starts with a 12-byte-length header and contains, at least, one DL-MAP-IE for each burst, which is transmitted in the DL subframe. As represented in Figure 2.2, each IE contains 44 bits of burst information and may contain one or more 16-bit-length CIDs based on the number of active connections in DL. The overall DL-MAP overhead, in bits, can be written as

$$\mathcal{O}_{\text{MAP}}^{\text{DL}} = 96 + (N_{\text{Burst}} \cdot 44) + (N_{\text{ActUsers}} \cdot 16) \quad \text{bits}, \tag{2.2}$$

where N_{Burst} is the number of data-bursts in DL subframe, and N_{ActUsers} is the number of users that are actively communicating with BS at given time instance.

As in UL each user usually transmits in a single burst, the number of bursts

in UL is equal to the number of active users. Therefore, the imposed overhead by UL-MAP that starts with a 7 byte header (56 bits), and carries UL-MAP-IEs of 32 bits-length. Thus the corresponding overhead can be estimated as:

$$\mathcal{O}_{\mathrm{MAP}}^{\mathrm{UL}} = 56 + (N_{\mathrm{ActUsers}} \cdot 32) \text{ bits} \tag{2.3}$$

MAP messages are transmitted with most robust MCS level (QPSK $\frac{1}{2}$) to be received with all users located in the service area. For more reliability, a repetition index is also applied over MAPs specified in FCH. Assuming $\times 4$ repetition, once the MAP messages overhead ($\mathcal{O}_{\mathrm{MAP}} = \mathcal{O}_{\mathrm{MAP}}^{\mathrm{DL}} + \mathcal{O}_{\mathrm{MAP}}^{\mathrm{UL}}$) is calculated, it is enlarged to closet multiple of 24 and then is multiplied by 4. This is due to the fact that, repetition operates on slots and not on bits.

The Channel Descriptor messages –i.e.: Downlink Channel Descriptor (DCD) and Uplink Channel Descriptor (UCD)– have been represented by using dashed lines in Figure 2.2. It means that these massages may not appear in all frames. Basically, the Channel Descriptor messages are sent periodically whenever the linking conditions change and new channel profile must be introduced to MSs. In our algorithm it is assumed that DCD/UCD messages appear every 100 frames. This results in 1.17 kbps of overhead for DCD and 1.71 kbps for UCD, assuming 7 burst-profiles. In addition to overhead introduced with control messages, sometimes, in PHY level the amount of data to be sent in a burst just spills over a slot. In these cases nearly an empty slot is sent, introducing additional channel overhead. To consider this condition, the algorithm assumes a 50%$\times$ MAU mismatch error for each data burst.

From MAC layer point of view, as explained earlier in data allocation procedure, MAC-PDU construction presents a number of overhead data-bits per PDU. There are a 6-byte generic header and a 4-byte CRC per PDU. It is likely to have an additional 3-byte overhead for either the packetization or fragmentation subheaders. Therefore, each PDU presents 13 bytes overhead ($13 \times 8 = 104$ bits) in MAC level as shown in Expression 2.4, where represents the number of PDUs in i^{th} burst.

$$\mathcal{O}_{\mathrm{PDU}}^{\mathrm{MAC}} = \sum_{i=1}^{N_{\mathrm{Burst}}} (N_i^{\mathrm{PDU}} \cdot 104) \text{ bits} \tag{2.4}$$

Generally, the number of PDUs per burst (N_i^{PDU}, i indexing each burst) is variable and is a function of the transmitted packets size and the in-use burst construction method. The standard supports multiple users burst in DL direction as long as the users application delay requirements are met. The limitations regarding to burst construction methods are properly dimensioning user rectangular burst and maintaining a single MCS mode during the entire burst. There are several proposals available in the literature exploring the burst construction problem. Fixed burst mapping considers a predetermined number of burst per frame, while Raster technique [114] assigns a burst to each data packet using a slot by slot scanning method. The first method, although very simple, requires advance knowledge of packet distribution, which might not be practical where the frame resources are flexible. The

second method imposes an *extra overhead* in both MAC and PHY layers. A more efficient packet method to achieve burst mapping has been presented in [115] with respect to a minimal component of a DL-burst, which is defined as "x slots in the frequency domain ($x \geq 1$)$\times$ one slot in the time domain". A number of minimal components are combined in time domain according to the packet size, and the units formed in this way may also be aggregated to form the burst as long as the same MCS mode is applied over them. Therefore, one can expect to have one multi-user DL-burst made of the groups of slots that support each specific MCS level in DL transmission. This strategy is used in this paper and is further studied in MCS distribution analysis in next subsection.

With these considerations in mind, Expression (2.5a) summarizes the overhead imposed on DL subframe by both PHY and MAC layers. By removing the overhead symbols, the useful resources are obtained that can be assigned for data transmission in DL. The overhead bits associated with UL subframe are fairly simpler to calculate, as MAP messages are not present in UL. Expression (2.5b) denotes the overall UL overhead.

$$\begin{aligned} \mathcal{O}^{\mathrm{DL}} = \quad & \text{preamble} + \mathrm{FCH} + \mathcal{O}_{\mathrm{MAP}} + \\ & + \mathcal{O}_{\mathrm{DCD+UCD}} + \text{mismatch} + \mathcal{O}_{\mathrm{PDU}}^{\mathrm{MAC}} \quad , \text{bits} \end{aligned} \tag{2.5a}$$

$$\begin{aligned} \mathcal{O}^{UL} = \quad & \text{ranging} + \mathcal{O}_{\mathrm{ACK/NACK}} + \\ & \mathrm{CQICH} + \text{mismatch} + \mathcal{O}_{\mathrm{PDU}}^{\mathrm{MAC}} \ (\text{bits}) \end{aligned} \tag{2.5b}$$

In these expressions, a mismatch overhead is assumed to model the possible existence of empty slots, as will be explained in the following section.

Having the overhead symbols removed, the useful data symbols (N_{sym}) is calculated as in Expression (2.6).

$$N_{\mathrm{sym}} = N_{\mathrm{sym}}^{\mathrm{DL}} - N_{\mathrm{sym}}^{\mathrm{DL},\mathcal{O}}, \tag{2.6}$$

where $N_{\mathrm{sym}}^{\mathrm{DL}}$ is the number of symbols in downlink subframe with respect to DL:UL Ratio, as explained in Section 2.2.2 and, $N_{\mathrm{sym}}^{\mathrm{DL},\mathcal{O}} = \mathcal{O}_{\mathrm{bits}}^{\mathrm{DL}}/N_{\mathrm{SC}}^{\mathrm{DL}}$ is the number of overhead symbols, where $N_{\mathrm{SC}}^{\mathrm{DL}}$ is the number of data subcarriers for DL as in Table 2.2 for each BW, while using PUSC.

2.5.2 Modulation and Coding Schemes Distribution

We have mentioned that Mobile WiMAX is able to support different modulation and coding schemes (MCS). The Link Adaption procedure in Mobile WiMAX significantly increases the system throughput by matching the most efficient MCS to each burst based on the channel conditions feedback. The choice of the MCS level depends on the Signal to Noise plus Interference Ratio (SINR), also called C/I,

through the perceived Block Error Rate (BLER). The most efficient MCS mode that achieves a BLER larger than a threshold stated for given SNIR will be the one to be used. The C/I vs. BLER link-level curves are available in the literature [116]. Generally, the users that are closer to the BS have a better chance for having higher MCS levels (and thus higher channel-rates), while those edge-users that suffer from higher interference and path-loss attenuation (lower SNIR) are assigned the least efficient MSC. This is the reason why, for capacity analysis purposes, the knowledge of MCS distribution of the target cell is required.

WiMAX Forum has stated the link budget assumptions and cell configuration parameters for Mobile-WiMAX profile [147]. One way is to assume the MCS distribution regions as concentric circles around the BS [105]. In this reference the data-rate of each region is also addressed as a function of path-loss and experienced SINR. On the other hand, a detailed system-level simulator configuration will result in more effective MCS distribution map. Reference [104] summarizes the consideration for system level modeling of IEEE 802.16e based networks. Having the appropriate simulator tool at the desired frequency band, the DL-SNIR coverage map (i.e.: physical link-layer) of the service area (i.e.: covered by target BS) is obtained. In frequency selective environments, such as OFDM systems, each subcarrier faces a different channel response. Therefore, WiMAX Forum specifications recommend the Exponential Effective SNIR Mapping as the PHY-abstraction for WiMAX Systems. The PHY abstraction predicts the BLER with respect to effective SNIR values for each MSC mode. In this respect, [100] and [101] present a list of minimum SNR needed for each MCS at BER = 10^{-6}, for CC and CTC coding respectively. The min-SNR values for turbo coding are shown in Table 2.3, together with the weight assumed for each MCS distribution representing the coverage percentage of the corresponding MCS level out of the total area under cover. Note that in this document it is assumed that the subscribers are uniformly distributed in the coverage area.

Having the overhead symbols removed, an accurate approximation of the data-rate offered by the system is achievable independent of the traffic load. In other words, system's data-rate is a function of the number of users and their distribution inside the coverage area, but not the service type of the users. The useful data symbols, N_{sym} as obtained in Expression 2.6, provide a number of slots that are available to be assigned to the users. The total number of useful slots in DL subframe ($N_{\text{slot}}^{\text{DL}}$) can be calculated as

$$N_{\text{slot}}^{\text{DL}} = \left\lfloor \frac{N_{\text{sym}} \times N_{\text{SC}}^{\text{DL}}}{48 \times \text{FRP}} \right\rfloor \tag{2.7}$$

where $N_{\text{SC}}^{\text{DL}}$ is the number of data subcarriers and FRP is the index of frequency reuse pattern, which is 3 for (1,1,3) and 1 for (1,3,3) pattern. Note that the (1,1,3) frequency reuse pattern divides the cell's available BW over 3 sectors (1/3 BW per sector), while in (1,3,3) pattern one frequency band is available for each sector (BW per sector). In first pattern it is assumed that the BS sends the synchronization and control messages using the entire BW and the remaining resources are equally

divided between 3 sectors.

When the users are uniformly distributed in the service area (i.e.: covered by target BS) at each time instance, the useful slots in the DL can be assigned to multi-user bursts based on the MCS distribution in the serving cell. Note that in this assignment, an integer number of slots must be optimally appointed to each MCS level, while the burst construction limitations explained in previous subsection are satisfied. Therefore, the i^{th} multi-user burst will contain $\lfloor N_{\text{slot}}^{\text{DL}} \times W_i^{\text{MCS}} \rfloor$ slots, where W_i^{MCS} denotes the percentage of the serving area that is covered by the i^{th} MCS level. Accordingly, $\lfloor N_{\text{slot}}^{\text{DL}} \times W_i^{\text{MCS}} \rfloor \times \text{BL}_i$ bits are offered in this burst for data transmission, where BL_i is the block size of the i^{th} MCS level. The *real, offered data-rate* in the entire downlink subframe ($\Omega_{\text{TOT}}^{\text{DL}}$) is then calculated with the summation of bits that all data bursts carry, divided by the frame duration (T_{frame}), that is:

$$\Omega_{\text{TOT}}^{\text{DL}} = \frac{\left\lfloor N_{\text{slot}}^{\text{DL}} \times \sum_{i=1}^{N_{\text{Burst}}} \left(\text{BL}_i \times W_i^{\text{MCS}} \right) \right\rfloor}{T_{\text{frame}}} \tag{2.8}$$

Note in in Expression 2.8 that the calculated data-rate $\Omega_{\text{TOT}}^{\text{DL}}(N)$ is a function of the number of users. By incrementing the N, the overhead increases and the available resource decreases. Therefore, given the system parameters and MCS distribution, it provides the available resource metric for each number of supported users.

This procedure may impose a number of empty slots per burst, even using a fair allocation assumption, known as "Over Allocation phenomenon". We have partly compensated this by assuming 50% mismatch error in the overhead calculations. Furthermore, there may remain a number of unallocated slots that do not fall within the rectangular confines of any data burst, that do not take part in useful resource calculation procedure. In the example calculations that will be conducted in this paper, we will substitute the parameters of Expression 2.8 with those that are assumed in Table 2.3, although they can be customized for any possible reception conditions. In this table, the example values for W_i^{MCS} are chosen close to that of a sub-urban area.

2.6 Service delivery Model and Subscribers Resource Consumption

As mentioned earlier, one key element in capacity estimation analyzes is to define a service delivery model for the system under study. According to [75] the system load is the minimum amount of shared resources required to serve the users. However, in new generation broadband wireless systems, including WiMAX, the users may demand different applications, each with different QoS requirements. Therefore, on the one hand, the scheduler must be aware of the available system resources, while, on the other hand, the scheduler must have the knowledge of users' demand

and their reception conditions.

With this in mind, in this section we define a generic model that integrates the application classes with their distributions and minimum data-rate requirement. In the first subsection we discuss the data-rate support requirements when multi-user services are jointly supported in the system. The second subsection considers the parameters that are required to statistically model the users application profile. In the last part, we formulate the entire elements that characterize our proposed service-load model along with calculating the required parameters in the parallel process for resource calculation purposes.

2.6.1 Data-rate requirements

The application classes and their QoS requirements for WiMAX systems have been studied in Section 2.2.6. In order to guarantee a quality user experience, a key feature of each data service is its data-rate requirement. In other words, some data services need the system to guarantee that, at each given time, a specific number of bits out of the entire available resources will be allocated for their transmission. Regarding this (i.e.: the level of required guarantee for data-rate access), the services are classified into 3 categories: (*a*) Constant Bit Rate (CBR) with strict data-rate requirement (Minimum Reserved); (*b*) Variable Bit Rate (VBR), which is characterized with a strict part (Minimum Reserved) and a upper bound data-rate (Maximum Sustained); (*c*) Best Effort Services (BE), which are served using the remaining bandwidth when other services are supported. Based on this data-rate requirement classification, the total available bandwidth is divided into *guaranteed* and *non-guaranteed* partitions. Figure 2.4 represents this bandwidth partitioning scheme while Table 2.4 shows the relation of each data-rate specification with its corresponding application and service flow category.

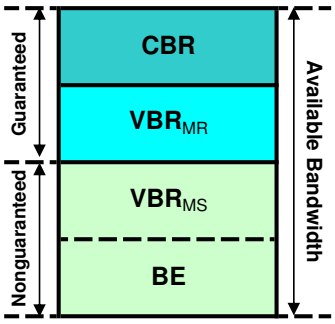


Figure 2.4: Simple representation of the bandwidth partitioning. Based on this data-rate requirement classification, the total available bandwidth is divided into *guaranteed* and *non-guaranteed* partitions. See the main text for further details.

The guaranteed bandwidth requires unconditional dedication of the resources that grant the needed data-rate. Table 2.5 lists the minimum guaranteed data-rate requirement for each application category introduced by WiMAX Forum, when applicable. On the other hand, the non-guaranteed partition must be fairly shared between requesting services, based on the scheduling policy. The combination of the minimum guaranteed and non-guaranteed resources allocated to the contenting users is calculated based on the subscribers application profile, as will be explained in the following subsection. In each step, a new user will be admitted *if there are enough free resources to satisfy its minimum data-rate requirement.*

Table 2.4: Relation of each QoS data-rate specification with its corresponding application and service flow category in Mobile WiMAX [147].

Service Flow Category	**Applications**	**QoS Data Rate Specification**
Unsolicated Grant Service (UGS)	VoIP	Sustained Rate (CBR)
Real-time Polling Service (rtPS)	Streaming Audio or Video	Minimum Reserved Rate Maximum Sustained Rate (VBR)
Extended rtPS (ErtPS)	Voice with Activity Detection	Minimum Reserved Rate Maximum Sustained Rate (VBR)
No Real Time Polling Service (nrtPS)	File Transfer Protocol (FTP)	Minimum Reserved Rate Maximum Sustained Rate (VBR,BE)
Best Effort Service (BE)	Data Transfer Web Surfing	Maximum Sustained Rate (BE)

2.6.2 Application profile and number of active users

In broadband wireless networks supporting multi-services, including WiMAX, an application usage profile is required as a consistent input for the purpose of evaluation of the dimensioning and capacitating studies. The application profile shall cover subscribers application usage trends, in which the two major metrics to be considered are the *applications penetration* and *contention ratio*. The penetration parameter represents the distribution of applications that most likely users will use, while the contention ratio is a measure of the simultaneity of the usage of each application.

The collective application distribution information of broadband mobile/wireless service subscribers are mapped into the WiMAX application classification model, in which we have assigned a weight to each class representing the average penetration

of each application among the subscribers. The third column of Table 2.5 illustrates the weight (W_j^{app}) corresponding to j^{th} application category, when averaging over different applications grouped into a category. In addition to distribution weights, the application profile should provide the information on relative active attempts of each application in each time instance that in this paper we call it application's contention ratio (CR).

In other words, if N_U is the number of the subscribers located inside the target cell, then $[N_U \times W_j^{\text{app}}]$ represents the number of subscribers that are likely to use the j^{th} application. In addition, if CR_j represents the contention ratio of the j^{th} application, at a given time, only $[N_U \times W_j^{\text{app}} \times \text{CR}_j]$ subscribers will be connected to the cell

In order to give reference values for the contention ratios of different applications, the subscribers are classified based on their end-user type. Regarding this, the users may subscribe for Residential or Business class services, each with different service profile. Statistical studies prove that the attempts per application per day are different for different end users [74]. According to global statistics of WiMAX deployments in [113], the percentage of users with residential profile is $N_R = 58\%$, while that of users with business class services is $N_B = 42\%$, $(N_R + N_B = 100\%)$

The detailed contention ratios for residential and business class subscribers are presented in [74] for each application per day. These ratios are averaged over the daily traffic trends and are shown in the corresponding column of Table 2.5 for each application category. Recalling the example in previous paragraph, the number of residential subscribers that are actively using the i^{th} application can be estimated as $[N_U \times N_R \times W_j^{\text{app}} \times \text{CR}_j^R]$, where CR_j^R is the average contention ratio for j^{th} application in residential class.

Thus, the number of active users N_{ActUsers} can be written as

$$N_{\text{ActUsers}} = N_U \times \left\lceil N_R \times \sum_{j=1}^{5} \left(W_j^{\text{app}} \times \text{CR}_j^R\right) + N_B \times \sum_{j=1}^{5} \left(W_j^{\text{app}} \times \text{CR}_j^B\right) \right\rceil, \quad (2.9)$$

where N_R and N_B are the percentage of the end users with Residential and Business profile respectively, and CR_j^R and CR_j^B are the average contention ratios for the j application in the Residential and Business profile respectively. The obtained N_{ActUsers} parameter here is also used in the parallel process for overhead estimation calculations.

Note that the values presented in Table 2.5 are for general reference and can be easily customized for each traffic scenario. In other words, the average number of attempts per active user during a day can be converted into the number of application sessions during a specific time of the day with diurnal-application distribution [74]. For instance, substituting the peak hour values instead of diurnal averaged ones may alter the application profile for the worst-case load profile.

Table 2.5: Service Delivery Model. Value of the different parameters used in the applied profiles as a function of the application class.

j	**Application Class**	**Guaranteed BW (kbps) [4]**	**Weight**	**Ave. Contention Ratio** [74]	
				Residential	**Business**
1	Multiplayer Interactive Gaiming	50	0.25	0.25	0.5
2	VoIP and Video Conference	32	0.1	0.6	0.75
3	Streaming Media	64	0.125	0.75	1
4	Web Browsing & Instant Message	Nominal	0.325	0.75	0.75
5	Media Content Downloading	BE	0.2	0.4	0.9

2.6.3 Subscribers Service Demand Calculation

Considering the Service delivery model presented in Table 2.5, the data-rate requirement for the instantaneous load imposed on the system is achievable. We assume that all the active users are scheduled for transmission in each DL subframe. The application classes with specified guaranteed data-rate (the first 3 application classes in Table 2.5) are separated from those with non-guaranteed data-rate. The number of subscribers in the first group N^{G}_{ActUsers} is obtained by solving the Expression 2.9 for the first 3 applications. The aggregated guaranteed data-rate (R_G) is then calculated using Expression 2.10.

$$R_G = N^{G}_{\text{ActUsers}} \times \sum_{j=1}^{3} \text{DR}_j. \tag{2.10}$$

where, $N^{G}_{\text{ActUsers}} = N \times \left\lceil N_R \times \sum_{j=1}^{3} \left(W_j^{\text{app}} \times \text{CR}_j^{R}\right) + N_B \times \sum_{j=1}^{3} \left(W_j^{\text{app}} \times \text{CR}_j^{B}\right) \right\rceil$, and DR_j is the guaranteed data-rate of the j^{th} application as in Table 2.5 ($j = 1, .., 3$).

At each given time, the remaining system bandwidth can be allocated to non-guaranteed services with respect to users service-rate parameter as determined by the subscription contract. Each subscriber is due to have access to a non-guaranteed data-rate up to the subscribed service-rate, as long as the existing higher priority services for that user (here the guaranteed data-rate services) are already served. The service-rate is an Application Profile input parameter and may be different for different end-user types. Let us assume that each residential subscriber may contract C_R kbps of service, while the business class subscriber's demand is C_B kbps. Therefore, we can introduce a measure for the maximum amount of aggregated data-rate for non-guaranteed service ($R^{\text{MAX}}_{\text{Non,G}}$) as conducted in Expression 2.11.

$$R^{\mathrm{MAX}}_{\mathrm{Non,G}} = N^{\mathrm{Non,G}}_{\mathrm{Active}} \times \left[N_R \times \left[C_R - \sum_{j=1}^{3} \mathrm{DR}_j \right] + N_B \times \left[C_B - \sum_{j=1}^{3} \mathrm{DR}_j \right] \right] \quad (2.11)$$

where $N^{\mathrm{Non,G}}_{\mathrm{Active}} = N_{\mathrm{ActUsers}} - N^{G}_{\mathrm{ActUsers}}$ is the number of active subscribers using applications with non-guaranteed data-rate, N_R and N_B are the percentage of the end users with Residential and Business profile respectively.

The total aggregated data-rate at each time instance is simply calculated by the summation of guaranteed and non-guaranteed data-rates (i.e.:$R^{DL}_{Tot} = R_G + R_{Non,G}$). Note that $R^{DL}_{Tot}(N)$ is a function of the number of users and linearly increases as the number of users increments. Given the application profile, $R^{DL}_{Tot}(N)$ provides the resource consumption metric for each number of supported users. Following the generic algorithm proposed in this paper, at each given time, the resource consumption $R^{DL}_{Tot}(N)$ for each number of users is calculated independent of available resources $\Omega^{\mathrm{DL}}_{\mathrm{TOT}}(N)$, that was calculated in Expression 2.9. For each number of users these two values are compared and if $\Omega^{\mathrm{DL}}_{\mathrm{TOT}}(N) \geq R^{DL}_{Tot}(N)$, N will be incremented by one. The largest N that satisfies the above condition is introduced as the system capacity in supporting multi-services.

Recalling the parallel process for overhead calculations, apart from N_{ActUsers} parameter, an estimation for the number of PDUs per burst (N^{PDU}_i) was also required as in Expression 2.4. Having the load per frame and the number of active users, a reasonable approximation for the average-packet-size (P_{Ave}) is calculated for each number of users. Assuming k averaged-size-packets per PDU $\left(i.e. : \mathrm{ARQ} - \mathrm{BLOCK} - \mathrm{SIZE} = k \times T_{\mathrm{frame}} \times \left(\Omega^{\mathrm{DL}}_{\mathrm{TOT}}/N_{\mathrm{ActUsers}}\right)\right)$, each burst will accommodate a number PDUs proportional to its MCS distribution weight. This relation is shown in Expression 2.12.

$$N^{\mathrm{PDU}}_i = \left\lceil N_{\mathrm{ActUsers}} \times \frac{W^{\mathrm{MCS}}_i}{k} \right\rceil . \quad (2.12)$$

Note that the calculated average packet size is not a practical concept, since the packets generated by the users considerably vary in size for different applications and also over time. However, this concept gives a good approximation for estimation of the number of PDUs per data burst by introducing a trade-off between different packet sizes.

2.7 Experimental work

This section aims at exploring the feasibility of the proposed capacity estimation methodology. We have carried out two groups of simulations "Experiment I" and "Experiment II", which will be described and discussed in Subsections 2.7.1 and 2.7.2, respectively.

2.7.1 Experiment I

This first experiment explores the general behavior of the proposed capacity estimation algorithm. This simulation requires two types of input parameters. Table 2.6 lists the two input parameter categories: "System Parameters" and "Subscriber Service Parameters". The system parameters that we have used are just the default values recommended by WiMAX Forum system evaluation methodology. They are also those common values used in practice. The subscription service parameters are based on the statistical data we have explained in the presented application profile.

Table 2.6: Input Parameters of Experiment I. The "system parameters" are the default values recommended by WiMAX Forum system evaluation methodology and also are the common values used in practice. The "subscription service parameters" are based on the statistical data, as explained in the presented application profile.

System Parameters	
Bandwidth (MHz)	10
DL:UL Ratio	2:1
Frequency Reuse Pattern (FRP)	1,3,3
MCS Distribution	Table-2
Subscriber Service Parameters	
N_R	58%
N_B	42%
C_R (kbps)	512
C_B (kbps)	1000
Load Model and Application Profile	Table-4
Simulation Results	
Max No. of supported multi-service users	24
NActive	16
Max goodput (Mbps at N=1)	8.51
Min No. of overhead slots (at N=1)	79
Sector Throughput (Mbps at N=24)	5.62
Max No. of overhead slots (at N=24)	279
Average-Packet-Size (bytes)	211

The numerical simulation results provided in Table 2.6 for Experiment I have been obtained with following conditions: (a) All the active services, including guaranteed and non-guaranteed data-rate services, are scheduled in the frame of the sector under study at each given time: (b) Non-guaranteed services are assigned a data-rate, equal to their service-rate subscription (refer to Section 2.6.3). The first assumption assures that the delay requirements of all the services are met and the

second one provides the highest user satisfaction with the price of imposing heavy load on the system.

With these considerations in mind, we can explore now the dynamic performance of the proposed algorithm. Just in this respect, Figure 2.5 represents the minimum service-load and the DL optimal goodput (kbps) as a function of active users in Experiment I. As shown, the resource consumption exceeds the available resource for the active user number 25. In other words, the two graphs representing the linear minimum service-load and the DL optimal goodput cross before meeting the data-rate requirement of 25th user. Thus, the specified sector can support 24 multi-service users with the sector goodput of 5.62 Mbps. By providing reference measures of magnitude, these results can help network planners to study the dimensioning capabilities of the implemented system.

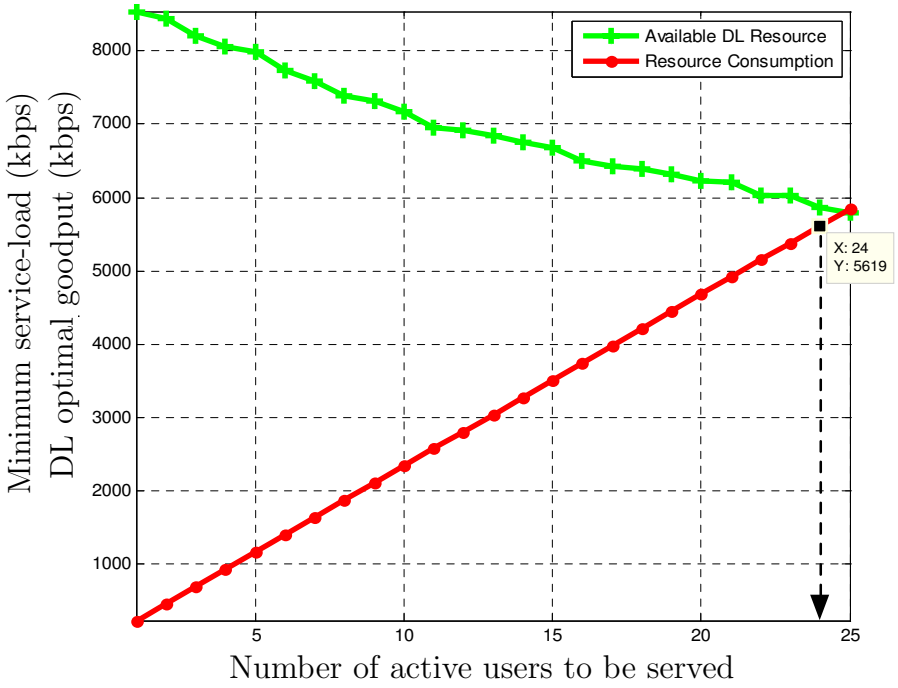


Figure 2.5: Dynamic trend of the algorithm in Experiment I: DL optimal goodput (green curve) and minimum service-load (red curve), both in kbps, as a function of the number of users to be served.

Figure 2.6-a shows the dynamic PHY+MAC overhead increase from above 3 symbols with single user up to 11 symbols in full-load. The un-integer values of symbols denote the overhead slots (i.e.: multiples of 24 bits in DL). The significant difference in number of overhead symbols in Figure 2.6-a, underlines the need to

consider a dynamic strategy for overhead determination with respect to the number of scheduled users in Mobile WiMAX capacity estimation analysis. The number of unallocated slots during the resource allocation procedure in Experiment-I is illustrated in Figure 2.6-b. These wasted resources are caused by burst construction limitations explained in the text. A considerable number of realizations show that an average of 3 slots remain unallocated in multi-burst mapping for mixed user traffic scenarios, while having a fair scheduling assumption.

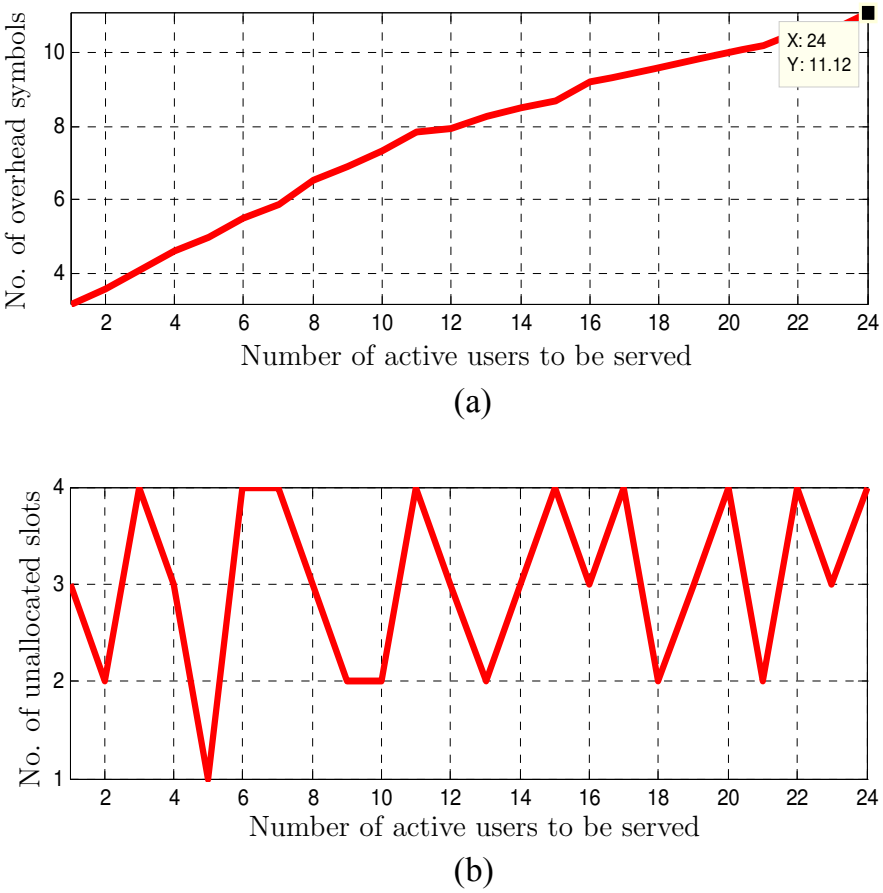


Figure 2.6: (a) Number of overhead symbols as a function of the number of active users to be served. (b) Number of overhead unallocated slots as a function of the number of active users to be served.

2.7.2 Experiment II

In Experiment II, we compare two different service provision scenarios for non-guaranteed services. The resource consumption in both scenarios are assumed the same. In this experiment, the input parameters for both system and subscription parameters are the same as Experiment I.

We model "Case-1", where the non-guaranteed services are using 50% of their maximum service-rate, and compare it with "Case-2" when 50% of the non-guaranteed services are not scheduled in target subframe, while the existing ones are using the maximum data-rate. In this way, the sector's load (subframe resource consumption) for both cases will be equal, but the sector's throughput and, hence, the number of supported users will be different. The simulation results show that the first scenario will support 32 multi-service users, while, in the second one, this value increases up to 38 users.

This is because of the fact that, by discarding 50% of the users with non-guaranteed data-rate, the overhead related to their signaling and controlling messages are also removed from available downlink subframe resources. Therefore, when the number of supported users is 32, the level of available resources in Case 2 is higher than Case 1.

As can be seen in Figure 2.7, the freed resources are enough to accommodate the $38 - 32 = 6$ extra multi-service users, with the specified application profile.

Note that the extra users also introduce additional overhead to the system. This can be noticed by the continuously declining slope of the available resources in Case 2. In spite of this, the simulation results prove that the released available resources in Case 2 can support a greater number of users up to the consumption of 5457 kbps. The specified system can support up to 106 users, when the entire available resources are assigned to the services with guaranteed data-rate. By providing reference measures of magnitude, these results can help the algorithm developers to evaluate and compare the service provision capabilities of different scheduling techniques.

The proposed algorithm is highly change sensitive and provides a simple environment to study the impact of each change (i.e.: system or load input parameters) on the system performance. Regarding this, and for the sake of completeness, we have repeated the experiments for different system parameters while the subscriber service parameters are the same as those listed in Table 2.6 and MCS distribution is as in Table 2.3. The results are presented in Table 2.7 for two different scheduling cases. The first one corresponds to the case in which the subframe is loaded with all services as in Experiment I. The second one corresponds to the case in which only guaranteed data-rate services are present and non-guaranteed services are not scheduled in the target subframe. It can be observed that the sector capacity can vary between 3 to 133 multi-users with different system parameters and subframe scheduling pattern.

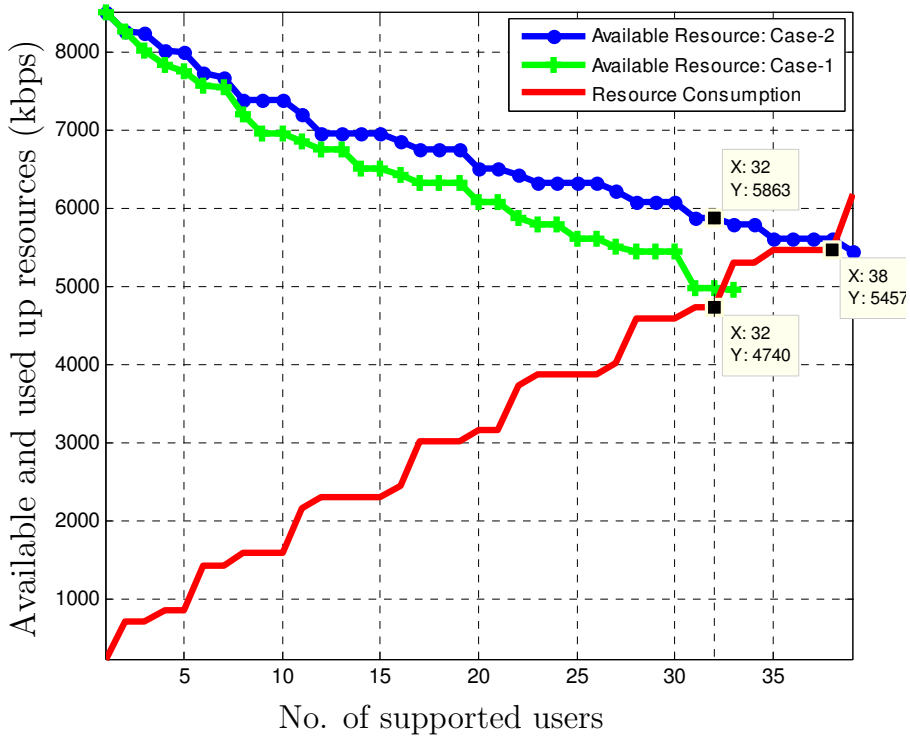


Figure 2.7: Resource consumption and available resources (kbps), in both "Case 1" and "Case 2", as a function of the number of supported users (Experiment II). See the main text for further details.

Table 2.7: Estimated capacity for different system parameters. The subscriber service parameters are the same as those listed in Table 2.6 and MCS distribution is the one in Table 2.3. "Gua." stands for guaranteed.

BW	**5 Mhz**				**10 Mhz**			
FRP	**1,1,3**		**1,3,3**		**1,1,3**		**1,3,3**	
DL:UL Scheduling	**Full**	**Gua.**	**Full**	**Gua.**	**Full**	**Gua.**	**Full**	**Gua.**
1:1	3	15	8	38	8	45	21	91
2:1	4	19	9	41	9	49	24	106
3:1	5	26	13	57	12	64	30	133

2.8 Summary and conclusions

In this chapter we have developed our dynamic capacity estimation methodology for Mobile WiMAX systems, whose main results have been published in [69]. Capacity refers in this chapter to the maximum number of multi-service users that can be simultaneously supported by a given access node.

We have explained why our algorithm is made of two major blocks, which, at each time, are solved in parallel to compute, on the one hand, the optimal available resources (considering the MCS distribution and multi-user burst construction), and to estimate, on the other hand, the minimum required resource consumption (with respect to application profile and the multi-service data-rate requirements).

We have also shown how properly computing the available resource requires to estimate first the necessary overhead at PHY + MAC levels. Our general formulations for overhead estimation can be used for any system parameter configuration.

The data-rate requirements of the covered multi-services have been grouped into guaranteed and non-guaranteed data-rates. We have also defined an application profile as a consistent input for capacitating and dimensioning studies evaluation.

We have carried out a number of experiments. Specifically, several simulations have been run with example values for MCS distribution and daily averaged application profile. The elaborated statistical application profile replaces the complexity of simulating a detailed scheduling strategy, thus providing reference measures in order to evaluate the performance of different scheduling techniques in real time.

The simulation results with regard to PHY+MAC overhead removal methodology underline the significant roll of overhead calculations in capacity estimation studies.

Having the number of supported user with each WiMAX hotspot, with regards to the parameters that characterize the access point and the service area, the network planners can optimize the service provision in the coverage area. Furthermore, the proposed algorithm is highly change sensitive. In other words, the algorithm can be simply configured for each set of input parameters, in order to study the impact of each parameter change on system performance. The performance is evaluated in a single output result, that is, the number of supported multi-users (or the cell goodput).

Further studies shall be done in order to improve the proposed algorithm with more implementation details, while keeping the simple structure of the proposed approach for capacity estimation. In continuation to this work, we plan to propose solutions in order to integrate the performance enhancing techniques in proposed algorithm and to include the short-scale channel state variations factor in resource allocation. Furthermore, the impacts of handover and mobility overheads on the Mobile WiMAX capacity estimation are other topics that will be discussed in our related future works.

Chapter 3

Overhead, Data Throughput and Capacity in LTE Downlink

3.1 Introduction

Long Term Evolution (LTE), the fastest-growing mobile technology [3,22], is able to provide unprecedented very high data rate and extremely low latency, specially in downlink (DL). This makes LTE very useful not only for conventional users [3] but also for others involving novel applications in machine-to-machine communication [28], mHealth services [29], smart grids [30], or green LTE [31, 32], even when potentially deployed in TV white spaces [33]. During Q3 2015 LTE has added around 120 million novel subscribers, while Wide-band Code Division Multiple Access (WCDMA) [12] have added only 70 million users, and even Global System for Mobile Communications (GSM) have declined by 60 million [3] –in spite of most users of Third and Four Generation (3G, 4G) use GSM and Enhanced Data Rates for GSM Evolution (EDGE) networks as a fallback [3]–. There are several causes explaining this vertiginous development. On one hand, user equipment (UE) manufacturers have made public an annual growth in 2015 of 79%, smartphones being 75% of all UEs and the strongest driving force for the aforementioned fast deployment. On the other hand, Orthogonal Frequency Division Multiple Access (OFDMA) technique in DL [34, 35] makes LTE able to support higher data rates than those of High-Speed Packet Access (HSPA) [36, 37]. Thus, LTE acceptance (up to 442 commercial LTE networks in 147 countries [3]) is driven by a strong demand for both better user experience and faster networks than those provided by HSPA technology [13]. All these combined factors are fueling an increasing demand of novel, more data-intensive services (such as TV, Video on Demand (VoD), gaming, etc.): data traffic has grown 65% between Q3 2014 and Q3 2015 [3]. The negative counterpart of the superior performance of LTE when compared to other mobile technologies is its highest complexity, which makes much more difficult its modeling, dimensioning [38] and planning [39–41], demanding new simulation and evaluation tools [42] along with novel algorithms for resource allocation [43] and handover in LTE macrocell-femtocell networks [44], to name a few.

Within this context, Figure 3.1 will assist us in introducing the purpose of this paper and in motivating its structure. It consists of three subfigures. Figure 3.1 (a) aims at illustrating the fact that the DL transmission resources in LTE exhibits dimensions of time, frequency and space. This latter, measured in "layers", is used by means of Multiple Antenna Technology (MAT), which involves several multi-antenna transmission and reception schemes. In this work we will use some concepts of these techniques, although the deepest specific details are beyond the scope of this paper. The interested reader is referred to Chapter 11 in [14] for further details. With this remark, Figure 3.1 (b) focuses only on one of such layers, and represents, in a simplified way, how OFDMA is used in the LTE DL, in the particular case of Frequency Division Multiple Access (FDD) and normal cyclic prefix. As illustrated, the basic unit for data transmission, called Physical Resource Block (PRB) or simply Resource Block (RB), is a two-dimensional (time, frequency) object in a grid. In time dimension, an RB last 0.5 ms, the length of 1 slot. Two slots form an 1-ms length subframe. A subframe is the smallest time interval in which LTE can assign RBs to UEs (or simply, users), and is called Transmission Time Interval (TTI)[1]. In the frequency dimension, an RB simply consists of 12 orthogonal subcarriers (SCs) assigned to an UE during one time slot, which in turn consists of a number of 7 OFDM symbols. The smallest information unit is the one carried by 1 SC during 1 time symbol (T_S), and is called Resource Element (RE). Most of the REs –white REs in Figure 3.1 (c)– are used to transmit *user data* on specialized physical channels such as Physical Downlink Shared Channels (PDSCHs). The rest of the REs, a reduced number of them, have to be reserved for a number of necessary network functionalities such as *signaling*, *control* or *synchronization* [30, 64–67] at both the Physical (PHY) level and Media Access Control (MAC) sub-level. These REs are "overhead" ($\mathcal{O}$) REs. Although it will be explained in detail later on, controlling the physical channels is motivated, for instance, by the need to decide, depending on user's Channel Quality Indicator (CQI), the Modulation and Coding Scheme (MCS) to be applied on each user's REs. Other overhead REs have to be reserved for Reference Signals (RSs) (dashed grey REs) –which transport some signaling information– and Synchronization Signals (SSs). The later has not been represented in Figure 3.1 (c) since, unlike the others, are *not* computed at TTI scale, but at *frame* scale.

As will be shown in the literature review in Section 3.2, overhead is usually estimated either as a constant value (regardless the operation conditions), or as a percentage of the total available REs, or assuming some overhead mechanisms as negligible, leading to overestimate bit rates or throughputs.

Another difficulty encountered when trying to analyze overhead is that most of the related information is spread over a number of recommendations, technical books

[1] Assigning resources to UEs every 1ms is a *dynamic* scheduling, which is efficient for services using the Transmission Control Protocol (TCP), in which the traffic is bursty and dynamic. In contrast, *persistent* scheduling –useful for services such as VoIP in which packets are small, periodic and semi-static in size– allows for allocating resources to UEs for a time period longer than one TTI [14].

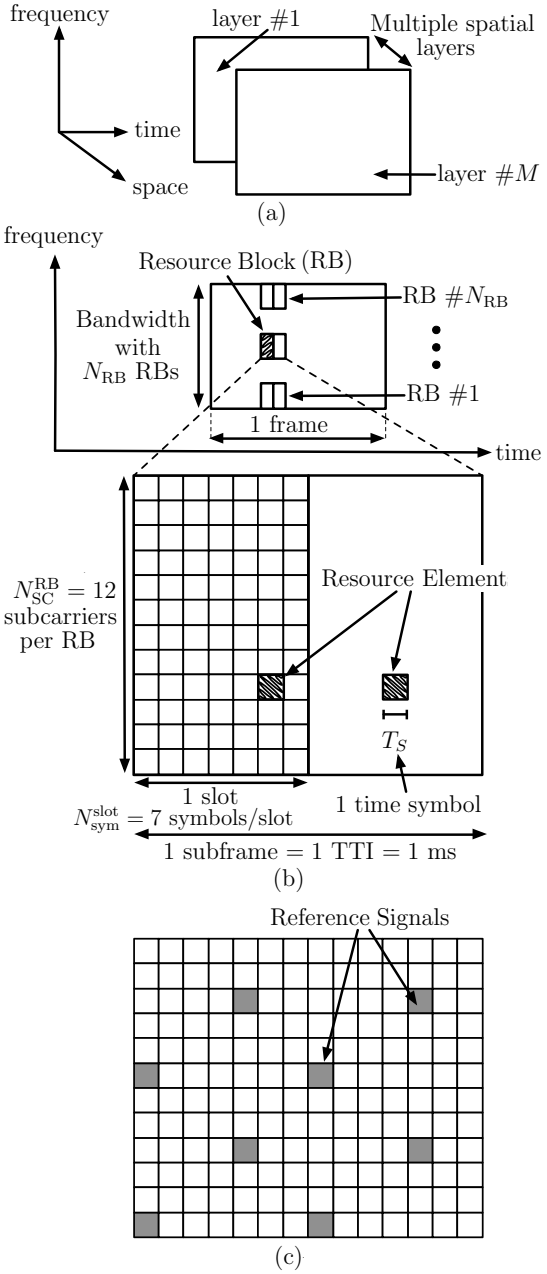


Figure 3.1: (a) Multiple spatial layer concept. (b) Time-frequency resource structure (normal cyclic prefix case). (c) Example of overhead. See the main text for further details.

or research papers, and to the best of our knowledge there is no paper modeling and discussing the influence that *all* the overhead mechanisms have on different aspects of LTE performance that will be shown later on.

Just in this respect, the *purpose* of this work is to collect overhead information and synthesize a *data-rate* based *performance model* that aims at accurately com-

puting all the overhead REs within a unified framework. Our *dynamic* overhead calculation methodology, which involves both concepts of the PHY level and MAC sub-level of the Data Link layer (Layer 2), takes into account all the possible variables it depends on: channel bandwidth, scheduling type, transmission antenna scheme, Hybrid Automatic Repeat Request (HARQ) configuration, and number and reception conditions of UEs. Therefore, the methodology, which can be easily adapted for different LTE system configurations and realistic situations, assists in: 1) accurately computing the contribution of *each* overhead mechanism and its corresponding influence on real data throughput; 2) optimizing throughput (by selecting, when possible, those configurations that minimize overhead); 3) calculating the maximum number of users per TTI (which are restricted by a limited number of controlling REs); and 4) dynamically computing throughputs, not only in single user scenarios, but also in multi-service and multi-user scenarios with adaptive multi-antenna transmission techniques for each user.

With this is mind, the rest of this paper is organized as follows. Section 3.2 reviews the related work, while Section 3.3 focuses on mathematically describing the problem and exploring the LTE standards and highlighting the technical concepts that are needed to characterize overhead. In this respect, we have aimed at reaching a balance between making this paper stand by itself and keeping the number of necessary technical details not too large. Using such concepts, Section 3.4 models the different overhead components while Section 3.5 centers on computing the total overhead at both TTI and frame scales. Sections 3.6 and 3.7 tackle, respectively, the problems of computing some required Initial Controlling Symbols REs and the maximum number of UEs that these enforce. Section 3.8 presents our methodology to accurately compute both the overhead and the data rate throughput in complex cases in which there are different users with difference services. Finally Sections 3.9 and 3.10 discuss the experimental work and the corresponding conclusions, respectively.

For the sake of clarity Table 3.1 lists the acronyms used in this paper.

3.2 Related work

Performance models can be classified into two basic categories. The first one, formed by *power-control* based models, focuses on power optimization by taking into account both the system configurations and the service area as a function of Signal-to-Interference-plus-Noise Ratio (SINR) distributions [40, 77]. Usually the involved powers are variables in optimizing objective functions (Channel Quality, Quality of Service –QoS–, etc.) which, in turn, are functions of SINR. There are several approaches for power optimization involving non-linear objective functions in which an interference matrix contains non-linear power coupling elements [78–81].

The second category of performance models, called *rate-control* models, tackles

Table 3.1: List of acronyms used in this paper.

Acronym	Meaning
2G, 3G, 4G	Second, Third, ForthGeneration
BS	Base Station
BW	Bandwidth
C-RNTI	Cell's Radio Network Temporary Identifier
CQI	Channel Quality Indicator
CRB	Contiguously allocated Resource Blocks
CRC	Cyclic Redundancy Check
DCI	Downlink Control Information
DL	Downlink
ECR	Effective Code Rate
EDGE	Enhanced Data Rates for GSM Evolution
eNB	Evolved Node B
FDD	Frequency Division Duplex
GSM	Global System for Mobile Communications
HARQ	Hybrid Automatic Repeat Request
HSPA	High Speed Packet Access
IP	Internet Protocol
LTE	Long Term Evolution
MAC	Media Access Control sub-level
MAT	Multiple Antenna Technology
MCS	Modulation and Coding Scheme
MIMO	Multiple-Input and Multiple-Output
OFDMA	Orthogonal Frequency-Division Multiple Access
PCFICH	Physical Control Format Indicator Channel
PDCCH	Physical Downlink Control Channel
PDSCH	Physical Downlink Shared Channels
PDU	Protocol Data Unit
PHICH	Physical Hybrid-ARQ Indicator Channel
PHY	Physical level
PRB	Physical Resource Block
QAM	Quadrature Amplitude Modulation
QoS	Quality of Service
QPSK	Quadrature phase-shift keying
RB	Resource Block
RBG	Resource Block Group
RE	Resource Element
ROHC	RObust Header Compression
RRC	Radio Resource Control
RS	Reference Signal
RTP	Real-time Transport Protocol
SC	Sub-carrier
SISO	Single-Input and Single-Output
SNIR	Signal to noise-and-interference ratio
SS	Synchronization Signal
TB	Transport Block
TBS	Transport Block Size
TCP	Transmission Control Protocol
TTI	Transmission Time Interval
TxD	Transmit Diversity
UDP	User Datagram Protocol
UE	User Equipment
VoD	Video on Demand
VoIP	Voice over IP
VRB	Virtual Resource Block
WCDMA	Wide-band Code Division Multiple Access

the problem of maximizing data-rate taking into account the combination of system configurations and data-rate provision as a function of the load (rate) distribution. The model we propose in this paper belongs to this rate-control approach. In contrast to power-control models, the service requirement in rate-control models is not an SINR threshold, but a required data volume to be served during a given time period [40, 41, 71]. The performance model is then the near-optimum solution to provide the demanded data-rate in the target area within the best system configuration. Although of great interest for LTE networks, however, rate-control models have been less investigated [30,66,67,72] since they have non-linear relations between cell loads [41].

Among the research papers on LTE data-rate performance models, overhead is usually estimated as either a constant value (regardless of operation conditions) [67], or as a percentage of the total available resources [30, 57, 66, 72], or even assuming

some overhead mechanisms as negligible [82]. Regarding this, the data-rate model in [82] aims at evaluating LTE performance, based on a clear overview of the basic LTE features, including its different transmission schemes. However, due to its introductory character, the fundamental concept of resource allocation in LTE has *not* been analyzed in detail. As a consequence, the model does *not* consider that overhead depends on the vast number of possible configuration flexibilities.

As mentioned, one of the approaches found in the literature consists in assuming overhead as a constant value [67, 83]. Specifically, the authors in [67] have expanded the peak data-rate model with an overhead count for a single user, although assuming a constant number of overhead controlling symbols per frame. Specifically, the results described in this work correspond to 2 particular cases of 1×1 and 4×4 schemes (n-Tx $\times$ m-Rx) for the lowest and highest bandwidths and for a single user, concluding that the throughput gain is proportional to the bandwidth expansion. The dynamic overhead calculation and multi-user throughput scenario with adaptive multi-antenna transmission techniques per user have *not* been considered.

Another interesting group of data-rate based models are those that assume overhead as a percentage of the total amount of available resources [30, 57, 66, 72, 84]. Specifically, the data-rate performance model explored in [72] aims at estimating LTE peak data-rate by considering that the transmitted data volume is equal to the highest Transport Block Size (the one delivered by the MAC sub-level to the PHY level during 1 TTI) allowed for a single user. RS overhead has been estimated as a percentage of the total amount of RE: 4.8%, 9.5% and 14.3% for 1,2, and 4 transmitting antennas respectively [72]. In a similar approach, the influence of overhead on data rate has also been studied in [66], which focuses on describing a system simulator that includes overhead for signaling and controlling as percentages of the total resources. In particular, in DL, RS overhead have been assumed to be 4.55%, 9.09%, and 12.12% for 1, 2, and 4 receiving antenna, respectively, while control overhead is assumed to be 1-3 symbols per subframe (about 12-21%). A third and more recent work is [30], which explores the feasibility of using LTE for smart grids. It makes use of a model that takes into account the number of REs available for the PDCCH (within a downlink sub-frame) by subtracting those REs reserved for Physical Control Format Indicator Channel (PCFICH), Physical Hybrid-ARQ Indicator Channel (PHICH), and RSs. However, SSs, at frame scale, have *not* been used in this work. Also in the approach of modeling overhead as a percentage of total resources, [84] has recently investigated the performance of signaling overhead caused by keep-alive messages (generated from always-on applications) based on LTE Radio Resource Control (RRC) state transition method for various RRC inactivity timer settings, although without considering the handover related signaling.

A key point to note in the works reviewed so far is that overhead, as we understand it, refers to those REs that are used to carry out signaling, control and synchronization information which involve tasks at PHY level and MAC sub-level. There are however other research papers that, using a number of different viewpoints, tackle particular aspects related to overhead, such as in tracking and paging [85–87], handover [88, 89], non-conventional CQI reporting schemes with reduced overhead [90],

core network signaling overhead [91], and those features related to overhead caused by protocols' *header information* in the levels above MAC sublayer [83, 92]. Regarding this miscellanea of heterogeneous papers with different approaches, all the overhead-headers added by protocols in layers above the MAC sublayer will be assumed in our approach as *part of the protocol data unit* (PDU) that the MAC sublayer receives. Examples of protocols (above the MAC sublayer) very commonly used in LTE are: Hypertext Transfer Protocol (HTTP) and Real-time Transport Protocol (RTP) in Layer 5 (Application), or Transmission Control Protocol (TCP) and User Datagram Protocol (UDP) in Layer 4 (Transport), or Internet Protocol (IP) in Layer 3 (Network), or RObust Header Compression (ROHC) protocol in Layer 2. For instance, Voice over IP (VoIP) in LTE uses the ROHC protocol for IP/UDP/RTP header-compression, and a semi-persistent scheduling strategy to assign resources to users (adequate for services characterized by periodic and frequent transmission of small packets such as VoIP) [93–96]. TCP/IP protocols are more used for mobile Internet, for which dynamic scheduling (every 1ms TTI) is more used [92]. Regardless of the protocols used for one service or another (and whose overhead-headers are already included in the PDU that the MAC sublayer receives), the overhead at PHY+MAC level described in Section 3.1 *must always be applied*.

Thus, this kind of overhead plays a key role to accurately compute LTE performance (basically, data throughput $\mathcal{T}$) in a variety of feasible configurations, and despite of this, it has not received much interest in the technical literature to the point that, to the best of our knowledge, there is *no* model containing all components and dependencies. In view of the lack of an in-depth and comprehensive LTE DL data throughput-based performance model including the influence of all PHY + MAC -related overhead contributions, we present the model that follows.

3.3 Problem statement: mathematical formulation and necessary technical background

3.3.1 Throughput and overhead: mathematical approach

As mentioned, one of the goals of this work is to accurately compute the DL data rate or throughput, $\mathcal{T}$ (bit/s), taking into account the existence of the mentioned overhead. $\mathcal{T}$ (bit/s) is simply the number of *physical-level bits* which are transmitted during one TTI, the aforementioned *minimum* interval time in which the system assigns resources to users. The user data block that the MAC sub-level delivers to the PHY one during 1 TTI is referred to as Transport Block (TB), while its corresponding number of physical-level bits is known Transport Block Size (TBS). Using such a terminology, the user throughput can be written as

$$\mathcal{T}(\text{bit/s}) = \frac{\text{TBS (bit)}}{\text{TTI (s)}} \Rightarrow \mathcal{T}(\text{bit/s}) = \frac{\text{TBS (bit)}}{10^{-3}\text{s}} = 10^3 \cdot \text{TBS(bit)}. \tag{3.1}$$

Intuitively, as can be inferred from what was mentioned in Section 3.1, user throughput increases as the number of assigned RBs (and consequently, the number of REs) gets bigger. As there are some REs (for instance, those necessary for Synchronization Signals) which are used *only once per frame*, throughput calculation may have *different* values depending on whether the time interval considered (Δt) is either 1 TTI or 1 frame. This is why we will label the *total* number of *useful* REs (that is, those for user data) during a time interval Δt (either 1 TTI or 1 frame) as

$$\mathcal{U}_{\mathrm{TOT}}^{\Delta t}\ [\mathrm{REs}] \doteq \mathcal{A}_{\mathrm{TOT}}^{\Delta t} - \mathcal{O}_{\mathrm{TOT}}^{\Delta t}, \tag{3.2}$$

where $\mathcal{A}_{\mathrm{TOT}}^{\Delta t}$ and $\mathcal{O}_{\mathrm{TOT}}^{\Delta t}$ represent, respectively, the total number of *available* (*all* the REs in Figure 3.1 (c)) and *overhead* REs. Superscript Δt in Expression (3.2) is not superfluous since, as will be shown later on, it will assist us in modeling a variety of different situations using a *unified formalism*. For instance, computing overhead at TTI scale ($\Delta t = 1$ TTI) is specially practical in applications using Transmission Control Protocol (TCP), in which user's traffic is bursty and dynamic. However, services such as VoIP, in which packets are small, periodic and semi-static in size, require persistent scheduling for allocating resources to a UE for a longer time period than one TTI ($\Delta t = p$ TTI, p being an integer).

To proceed further with LTE throughput modeling, it is convenient to consider that *not* all the *physical-level bits* are used to transport user data but also to detect errors –for instance, Cyclic Redundancy Check (CRC) bits–. The number of CRC bits to be appended to a TB is related to the channel coding applied, which, in sort, consists in [14, 148]: 1) Appending 24 CRC bits (to provide error detection) at the end of any TB; 2) Segmenting the transport block into code blocks; 3) Processing each code block with a 1/3 turbo coder; 4) Reassembling the resulting code blocks forming a single *codeword*.

The practical consequence for the purpose of this paper is that, essentially, a codeword is a TB with error protection (1 codeword = TBS +24 bits), whose 24 CRC-bits can *not* been used for user data. This affects the *efficiency* of this channel coding process. Such efficiency can be characterized by using the Effective Code Rate (ECR) concept. It can be defined as

$$\mathrm{ECR} = \frac{\mathrm{TBS} + 24}{\mathcal{U}_{\mathrm{TOT}}^{\mathrm{TTI}} \cdot Q} \tag{3.3}$$

where Q is the *modulation order* or number of bits associated to the MCS used.

The Effective Code Rate in Expression (3.3) can only take a number of values in LTE. These values have been defined in the Technical Specification [149] as a function of the *channel quality* the UE is receiving. This quality is quantized by a parameter called Channel Quality Indicator (CQI) [149], directly related to the Signal to noise-and-interference ratio (SNIR) at user's receptor. Table 3.2 lists the ECR values used in LTE (labeled $\mathrm{ECR}^{\mathrm{MCS}}$) as a function of the CQI values, which have been defined as integer numbers ranging from 1 to 15. Note that high CQI values ($10 \leqslant \mathrm{CQI} \leqslant 15$) mean that the UE reception is very good, what allows for

using high order modulation schemes (64-QAM, in this CQI interval).

Table 3.2: Effective Code Rate values used in LTE (labeled ECR^{MCS}) as a function of the Channel Quality Indicator (CQI). MCS stands for Modulation and Coding Scheme. Q is the modulation order. Information extracted from [149].

CQI	Modulation	Q	ECR^{MCS}
1	QPSK	2	0.076
2			0.117
3			0.188
4			0.301
5			0.438
6			0.588
7	16QAM	4	0.369
8			0.478
9			0.601
10	64QAM	6	0.455
11			0.554
12			0.650
13			0.754
14			0.825
15			0.926

Accordingly, considering the allowed ECR^{MCS} values in Table 3.2, the Transport Block Size in Expression (3.3) can be rewritten as

$$\text{TBS [bits]} = \text{ECR}^{\text{MCS}} \cdot \mathcal{U}_{\text{TOT}}^{\text{TTI}} \cdot Q - 24, \tag{3.4}$$

whose value can be increased by using multiple antenna transmission schemes [14]. Specifically, to increase data rate, multiple antennas can be used: 1) in Transmitting Diversity (TxD) –to increase of the transmission robustness against multipath fading– or 2) in Spatial Multiplexing –also called MIMO (Multiple-Input and Multiple-Output) [150], which transmits multiple signal streams to a single user on multiple spatial layers created by antennas combinations–. Thus, there are *several* parallel *streams* between the n transmitting (Tx) antennae and the m receiving (Rx) antennae over the same bandwidth [60, 151, 152]. Thus, when using multiple antennae schemes, Expression (3.4) becomes into

$$\text{TBS} = \Omega \cdot (\text{ECR}^{\text{MCS}} \cdot \mathcal{U}_{\text{TOT}}^{\text{TTI}} \cdot Q - 24), \tag{3.5}$$

where Ω is a multi-antenna transmission factor related to the *maximum number of data streams* that can be transmitted between the involved antennae.

Note that Expression (3.5) represents the TBS (bits) in the simplified case in which: 1) There is only one user so that *all* the available data resources can be assigned to it; and 2) All the useful resource elements $\mathcal{U}_{\text{TOT}}^{\text{TTI}}$ use the same MCS.

Regarding this, we can remove the "single-user hypothesis" by assuming that there are $N_{\text{UE}}^{\text{TTI}_i}$ users to be served during a generic TTI_i. Depending on the schedul-

ing strategy adopted [153–155], a fraction w_j of the RBs could be assigned to each j (with $j = 1, 2, \cdots, N_{\mathrm{UE}}^{\mathrm{TTI}_i}$) during TTI_i,

$$w_j = \frac{N_{\mathrm{RB}}^{\mathrm{UE}_j}}{N_{\mathrm{RB}}} \tag{3.6}$$

$N_{\mathrm{RB}}^{\mathrm{UE}_j}$ being the number of RBs assigned to user j. Note that Expression (3.6) must fulfill

$$\sum_{j=1}^{N_{\mathrm{UE}}^{\mathrm{TTI}_i}} w_j = 1. \tag{3.7}$$

With the condition stated by Expression (3.7), the number of bits in the TBS given by Expression (3.5) leads to

$$\mathrm{TBS}_i = \sum_{j=1}^{N_{\mathrm{UE}}^{\mathrm{TTI}_i}} \Omega_j \cdot \left(w_j \cdot \mathcal{U}_{\mathrm{TOT}}^{\mathrm{TTI}_i} \cdot Q_j \cdot \mathrm{ECR}_j^{\mathrm{MCS}} - N_{\mathrm{CRC}}^i \cdot 24 \right) \tag{3.8}$$

where subscript i refers to TTI number i and subscript j has been included in Q_j and $\mathrm{ECR}_j^{\mathrm{MCS}}$ (because they can take different values for any user j). $\mathcal{U}_{\mathrm{TOT}}^{\mathrm{TTI}_i}$ is simply Expression (3.2) for $\Delta t = 1$ TTI, that is, $\mathcal{U}_{\mathrm{TOT}}^{\mathrm{TTI}_i} = \mathcal{A}_{\mathrm{TOT}}^{\mathrm{TTI}_i} - \mathcal{O}_{\mathrm{TOT}}^{\mathrm{TTI}_i}$. N_{CRC}^i is the number of CRC units per TB in TTI number i.

Assuming *full buffer traffic per user* (UEs are continuously downloading data at whatever data rate they can achieve) [156], then $N_{\mathrm{CRC}}^i = N_{\mathrm{UE}}^i$, and thus

$$\mathrm{TBS}_i = \sum_{j=1}^{N_{\mathrm{UE}}^{\mathrm{TTI}_i}} \Omega_j \cdot \left(w_j \cdot \mathcal{U}_{\mathrm{TOT}}^{\mathrm{TTI}_i} \cdot Q_j \cdot \mathrm{ECR}_j^{\mathrm{MCS}} - N_{\mathrm{UE}}^i \cdot 24 \right) \tag{3.9}$$

Finally, we can proceed further by removing the second simplifying hypothesis by considering the most realistic case in which each MIMO stream is transmitted with *different* MCS parameters. Accordingly, for each MIMO user with $n = 1, \cdots, \Omega_j$ (j indexing the users in TTI number i), Expression (3.9) can be generalized to

$$\mathrm{TBS}_i = \sum_{j=1}^{N_{\mathrm{UE}}^{\mathrm{TTI}_i}} \sum_{n=1}^{\Omega_j} \left(w_j \cdot \mathcal{U}_{\mathrm{TOT}}^{\mathrm{TTI}_i} \cdot Q_{j,n} \cdot \mathrm{ECR}_{j,n}^{\mathrm{MCS}} - N_{\mathrm{UE}}^i \cdot 24 \right) \tag{3.10}$$

Note that TBS increases with $\mathcal{U}_{\mathrm{TOT}}^{\mathrm{TTI}}$. As stated in Expression (3.2), for $\Delta t = 1$ TTI, the only feasible way for $\mathcal{U}_{\mathrm{TOT}}^{\mathrm{TTI}}$ to grow is that $\mathcal{O}_{\mathrm{TOT}}^{\mathrm{TTI}}$ decreases, since $\mathcal{A}_{\mathrm{TOT}}^{\mathrm{TTI}}$ is a constant value once the bandwidth to be used has been chosen. The reason is as follows. Since a RE is the grid element consisting in the use of an SC during a time symbol, then the number of available resource elements in a TTI (the scheduling time interval) will be

$$\mathcal{A}_{\mathrm{TOT}}^{\mathrm{TTI}} = N_{\mathrm{sym}}^{\mathrm{slot}} \cdot N_{\mathrm{slot}}^{\mathrm{TTI}} \cdot N_{\mathrm{SC}}^{\mathrm{RB}} \cdot N_{\mathrm{RB}}, \tag{3.11}$$

where $N_{\mathrm{sym}}^{\mathrm{slot}}$ is the number of symbols per slot, $N_{\mathrm{slot}}^{\mathrm{TTI}}$ is the number of slots per TTI, $N_{\mathrm{SC}}^{\mathrm{RB}}$ is the number of SC per RB, and N_{RB} is the number of RB. $N_{\mathrm{sym}}^{\mathrm{slot}}$ and N_{RB} depend on some details of the two-dimensional RE grid. In the time domain, $N_{\mathrm{sym}}^{\mathrm{slot}}$, is determined by the use of a Cyclic Prefix (CP) to mitigate the Inter Symbol Interference (ISI). Considering the normal CP duration, each slot can accommodate $N_{\mathrm{sym}}^{\mathrm{slot}} = 7$ OFDM symbols of 66.7 μs duration each. In the frequency domain, with a SC spacing of 15 KHz, the number of RBs to be assigned to a UE, N_{RB}, depends on the *scalable* channel bandwidth (BW): 1.4, 3, 5, 10, 15 and 20 MHz. Table 3.3 lists N_{RB} as a function of the channel bandwidth [157]. In turn, the resource blocks to be used are grouped in blocks called "Resource Block Groups" (RBGs), whose size P [157] has also been listed in Table 3.3. Parameter P is called "RBG size". For instance, if the channel bandwidth to be used is BW= 20 MHz then, according to Table 3.3, $N_{\mathrm{RB}} = 100$ and $P = 4$.

With these details in mind ($N_{\mathrm{sym}}^{\mathrm{slot}} = 7$, $N_{\mathrm{slot}}^{\mathrm{TTI}} = 2$ and $N_{\mathrm{SC}}^{\mathrm{RB}} = 12$), the number of available REs per TTI stated by (3.11) leads to

$$\mathcal{A}_{\mathrm{TOT}}^{\mathrm{TTI}} = 168 \cdot N_{\mathrm{RB}} \tag{3.12}$$

which, as mentioned, is a constant value once N_{RB} is set through the choice of BW in Table 3.3. This leads again to the conclusion that, to maximize TBS in (3.10), $\mathcal{O}_{\mathrm{TOT}}^{\Delta t}$ must be minimized. And to do that properly, we must model $\mathcal{O}_{\mathrm{TOT}}^{\Delta t}$, which will be done in Section 3.4. Prior to do this, we need some concepts, which will be summarized in the following Subsection 3.3.2.

Table 3.3: Number of resource blocks (N_{RB}) and Resource Block Groups size (P) as a function of LTE bandwidth (BW) [157].

BW (MHz)	1.4	3	5	10	15	20
N_{RB}	6	15	25	50	75	100
RGB Size (P)	1	2	2	3	4	4

3.3.2 Additional background

As mentioned before, most of the available REs $\mathcal{A}_{\mathrm{TOT}}^{\mathrm{TTI}}$ in Expression (3.12) – white REs in Figure 3.1 (c)– are used to transmit *user data*. The rest, the overhead REs, have to be reserved for a number of tasks related to *signaling*, *control* or

synchronization. Accurately computing $\mathcal{O}_{\mathrm{TOT}}^{\Delta t}$ (Section 3.4) demands some necessary background involving not only resource allocation (Subsection 3.3.2) but also some additional PHY layer processes and multi-antenna concepts (Subsection 3.3.2). These are the objectives of the following two sections.

Resource Allocation: some necessary details

We have mentioned that the smallest resource allocation unit in LTE is the PRB assigned during 1 TTI, the smallest scheduling time interval (1 ms subframe in Figure 3.1 (b)). That is, the user' TB is delivered by the MAC sublayer to the PHY layer (in a *group* of RBs) once every TTI. Aiming at assisting the DL scheduler to properly decide on resource allocation, the instantaneous channel-quality at any UE has to be estimated and sent back to the base station (e-Node B), ideally, as often as once per subframe [82]. In this respect, the Physical Downlink Control Channel (PDCCH) carries, for each UE, these scheduling assignments in a group of bits called "Downlink Control Information" (DCI) [149]. DCI carries, among others, two essential resource allocation data: 1) resource assignment information (the dimension of the resources assigned to each UE); and 2) the MCS to be used in the scheduled resources.

The dimension or amount of resources assigned to a given user is basically the number of PRBs to fulfill his/her communication needs. The required PRBs are grouped depending on a number of strategies. LTE standard [149] defines three DL "Resource Allocation Types" (Type 0, Type 1, and Type 2) [14]. Assuming that the PRBs are ordered in increasing frequency, as shown in Figure 3.2, these are mapped to the so-called "Virtual Resource Blocks" (VRB), labeled by using identifying numbers (VRB #). In this context, a Resource Block Group (RBG) is simply a set of consecutive VRBs. For illustrative purposes, Figure 3.2 represents an example with 13 RBGs (ordered with numbers $\#0, \cdots, \#12$) for a bandwidth BW = 5 MHz.

In general, the number of RBs in each RBG, the "RBG size" P (already listed in Table 3.3), is a function of the system bandwidth, BW. Thus, once the BW value to be used has been stated, according to Table 3.3, the number of RBs to be used, N_{RB}, is automatically set, so that the number of Resource Block Groups for DL will be [14, 158]

$$N_{\mathrm{RBG}} = \left\lceil \frac{N_{\mathrm{RB}}}{P} \right\rceil, \tag{3.13}$$

which, as will be shown, plays an important role when computing the overhead in an accurate way.

Figure 3.2 will also allow us to briefly describe the details of resource allocation types that play a role in computing overhead. The VRB numberings can correspond either to *localized* PRBs or to *distributed* PRBs in frequency space. Resource allocation Type 0 uses *localized* RBGs, as those represented in red color in Fig-

ure 3.2. Resource allocation Type 1 adds frequency diversity [14] to Type 0, by re-grouping the RBGs into "subsets". Each subset contains a selected number of VRBs distributed in frequency within *different* RBGs. Figure 3.2 shows two subsets ("Subset 0" and "Subset 1") in resource allocation Type 1. Finally, in resource allocation Type 2, the DCI resource assignment information indicates (to each assigned UE) a set of contiguously allocated VRBs, either localized or distributed. *Localized* VRB allocations can range from a single VRB up to a maximum number of VRBs constrained by the system bandwidth. In this case, as represented in Figure 3.2, the resource allocation field consists of a resource indication value labeled "starting resource block" (RB_{start}), and a length in terms of "Contiguously allocated Resource Blocks" (CRB), L_{CBR}. The reader interested in deeper details is referred to Section 7.1.6 in [149].

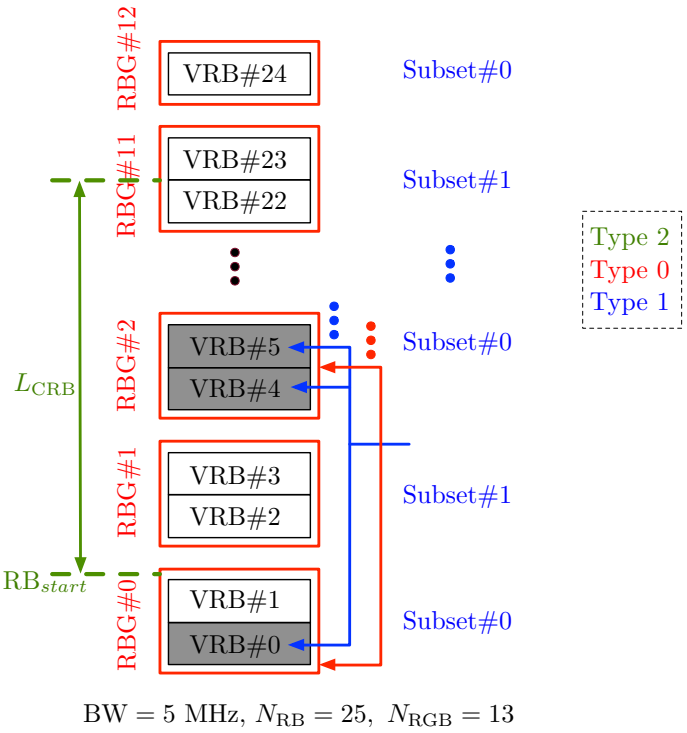


Figure 3.2: Simplified representation of an example of the Resource Allocation Types when BW= 5 MHz. See the main text for further details.

The DL Resource Allocation Types (Type 0, Type 1, and Type 2) are encoded in a predefined way using DCI formats, f_{DCI}. These depend on not only the way the PRBs and VRBs are grouped but also on the multi-antenna transmission scheme used. Including these concepts in our modeling is just the purpose of the following

subsection, and the reason why we postpone the necessary description of f_{DCI} to Table 3.5.

PHY Layer Processes and Multi-antenna Transmission

As introduced in Section 3.3, the user data payload (protocol data unit, PDU) delivered by the MAC sublayer to the PHY level is used to form the transmission codeword, which, as mentioned, can be seen as a TB with error protection, because of the 24-bit CRC attachment and turbo coding. As will be shown, the way the codewords are processed plays a key role in the overhead modeling.

Depending on the transmission mode to be used, up to 2 codewords can be transmitted in one subframe. The number of codewords is indicated by using the parameter q. It only takes two values: $q \in \{0, 1\}$, where $q = 0$ means single codeword, and $q = 1$ refers to 2 codewords. Using one value or another depends on the DCI format f_{DCI}, which, as will be summarized in Table 3.5, contains information about the resource scheduling type and about the multi-antenna transmission scheme. Note that it is necessary for any UE to receive information of these and other technical details. Figure 3.3 is useful in this respect because it represents schematically the PHY-layer processes involved [14, 159].

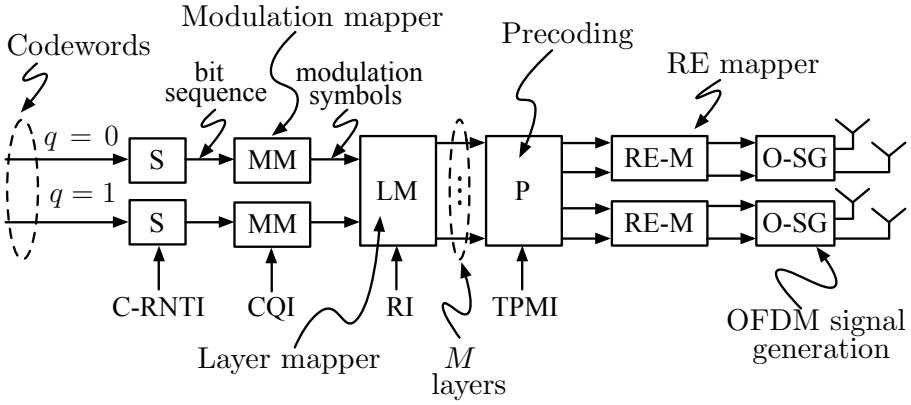


Figure 3.3: Simplified representation of Physical Channel Processing. See the main text for further details.

In the first step of Figure 3.3, the codewords are scrambled using an identifier called C-RNTI (Cell's Radio Network Temporary Identifier), generating a bit sequence, which is then modulated using the assigned modulation scheme (specified by the corresponding user's CQI value). The resulting *modulation symbols* for each codeword are processed by a Layer Mapper with M layers. The number of layers to be used, M, depend on the Rank Indication (RI) sent in the uplink from the UE. The Layer Mapper works as follows:

- In the case of a *single* antenna transmission, the content of the single codeword ($q = 0$) is simply mapped to a single layer.

- For *transmit diversity*, the standard [159] only allows the use of a single codeword ($q = 0$), and thus the modulation symbols of the codeword are distributed among 2 or 4 layers. Note that in this case, *transmit diversity* only improves the UE's reception condition by duplicating the same transmitted streams, thus allowing the use of higher order MCS.

- In *spatial multiplexing*, either one or two codeword can be utilized ($q = 0$ or $q = 1$, respectively). In spatial multiplexing, also known as MIMO, different parallel streams are transmitted over each antenna, thus increasing the data-rate two-fold. In MIMO implementations, generally, one codeword is mapped to one layer, while 2 codewords to 2 layers (2×2 MIMO).

In a configuration with up to 4×4 antenna scheme ($4Tx \times 4Rx$ transmit diversity or 4×4 MIMO spatial multiplexing), the available combinations form the 5 possible layer mapping cases, based on the specifications [159], are listed in Table 3.4.

Table 3.4: Summary of layer mapping cases for spatial multiplexing in a configuration with up to 4×4 MIMO, derived from [159]. It provides factor $\Omega_{\max}$ as a function of the layer mapping used, depending on the number of layers N.

Case	Number of codewords	Number of layers, N	Layer Mapping cases	$\Omega_{\max}$
A	1 ($q = 0$)	1	Single codeword mapped to single layer	1
B	1 ($q = 0$)	2	Codeword symbols are splitted (even/odd) between 2 layers. (Only used for 4TX)	1
C	2 ($q = 1$)	2	1^{st} codeword is mapped to the 1^{st} layer 2^{st} codeword is mapped to the 2^{st} layer (Both codewords are of the same length)	2
D	2 ($q = 1$)	3	1^{st} codeword is mapped to the 1^{st} layer 2^{st} codeword is splitted (even/odd) between the 2^{nd} and 3^{th} layers. (1^{st} codeword is half the length of the 2^{nd} one.)	2
E	2 ($q = 1$)	4	1^{st} codeword is splitted (even/odd) between the 1^{st} and 2^{nd} layers. 2^{nd} codeword is splitted (even/odd) between the 3^{rd} and 4^{th} layers. (Both codeword are of the same length.)	4

The multi-antenna transmission factor, Ω, on the rightmost column, and early introduced in Expression (3.5), is related to the number of streams that can be transmitted between the involved antenna [14]. In other words, Ω is the *Spectral Efficiency Gain*. The parameter $\Omega_{\max}$ represents the maximum multiplexing gain for each spatial multiplexing scheme that will be used for data-rate calculation purposes.

Once the layer mapping is carried out, the precoding stage in Figure 3.3 (which relates M layers to N antenna ports, with $M \leqslant N$) is applied to the modulation symbols of each layer. For doing this, the corresponding precoding information, the "Transmit Precoding Matrix Information" (TPMI), is communicated to the UE in the DCI, along with other resource allocation information.

Finally, the precoded symbols are mapped to the appropriate antenna ports, generating the corresponding OFDM signals. For further details, which would be out of the scope of this paper, the interested reader is referred to [150].

It is worth to mention that in MIMO configurations, the parallel streams are interleaved between the MIMO sub-channels. Thus, the average SINR experienced by the transmitted symbols is the average of the SINRs from all MIMO sub-channels. The theoretical 2-fold and 4-fold gains (i.e the $\Omega_{\max}$ values in Table 3.4) for 2×2 and 4×4 MIMO multiplexing, is conditional to the fact that the parallel channels maintain their SNIR high enough, so that all channels can support the same MCS level. The report in [160], summarizes the results of various 3GPP performance verifications and simulations studies on estimating the spectral efficiency conversion gains of different MIMO schemes. Considering Single Input and Single Output (SISO) antenna scheme as reference case (with $\Omega = 1$), this study reveals that the *Effective Spectral Efficiency Gain* (Ω) for spacial multiplexing using 2 Tx and 2 Rx antennas is 1.7, instead of 2, and for spacial multiplexing using 4 Tx and 4 Rx antennas is 3.2, instead of 4. In other words:

- for 2×2 MIMO: $\Omega_{\max} = 2$, and $\Omega = 1.7$
- for 4×4 MIMO: $\Omega_{\max} = 4$, and $\Omega = 3.2$

We complete this section by using Table 3.5, which summarizes those possible cases that combine the resource allocation strategies (Subsection 3.3.2) along with the multi-antenna transmission schemes (3.3.2) [60, 150]. Each combination corresponds to a DCI formats f_{DCI}. These and the number of codewords to be used (in terms of factor q) will assists us in accurately modeling the involved overheads. This is just the purpose of Section 3.4.

The reader is referred to [150] for further details on multi antenna techniques and precoding mechanisms in LTE.

Table 3.5: Downlink Resource Allocation Types (Type 0, Type 1, and Type 2), and the description of the DCI formats f_{DCI}. Open-loop MIMO does not require knowledge of the channel at the transmitter. Derived from [158].

DL Scheduling type	DCI format (f_{DCI})	Content	q
Type 0 (RGB based)	1	RGB based, single port + TxD	0
	2	RGB based, Closed Loop MIMO	0, 1
	2A	RGB based, Open Loop MIMO	0, 1
Type 1 (Subset based)	1	Subset Based, Single Port + TxD	0
	2	Subset Based, Closed Loop MIMO	0, 1
	2A	Subset Based, Open Loop MIMO	0, 1
Type 2	1	VRB Based, TXS	0
	2	VRB Based, Closed Loop MIMO	0
	2	VRB Compact	0
	2A	VRB Based, MU-MIMO [151] + Power Offset	0

3.4 Overhead mechanisms in Physical Layer for Downlink

There are several mechanisms that generate overhead: Physical Downlink Control Channel Overhead (3.4.1), Reference Signal Overhead (3.4.1), Physical Hybrid ARQ Indicator Channel Overhead (3.4.1), Physical Format Indicator Channel Overhead (3.4.1), Synchronization Signal Overhead (3.4.2) and Broadcast Channel Overhead (3.4.2). The last two are repeated only once per frame, while the first four are repeated once per TTI.

3.4.1 Overhead mechanisms with period 1 TTI

As reviewed earlier, a TTI is the minimum allocation unit in LTE. In addition to users data, the TTI shall carry the controlling and signaling information within Physical Downlink Shared Channel (PDSCH). The PDSCH occupies a number of initial symbols of each TTI, as illustrated in Figure 3.4, thus reducing the available space for data symbols. In continuation we review: the purpose for each of the overhead signals in PDSCH, how they are formed in LTE, and what overhead they impose per TTI.

Modeling the Physical Downlink Control Channel Overhead

We have mentioned in Subsection 3.3.2 that the DCI in the PDCCH carries UE-specific scheduling assignments. The DCI has a variable size (number of bits) because it carries different information for a wide variety of cases (as listed in Table 3.5). To identify these cases, we have also mentioned that the Technical Specification [149] defines different DCI formats, f_{DCI}, which are nothing but predefined formats

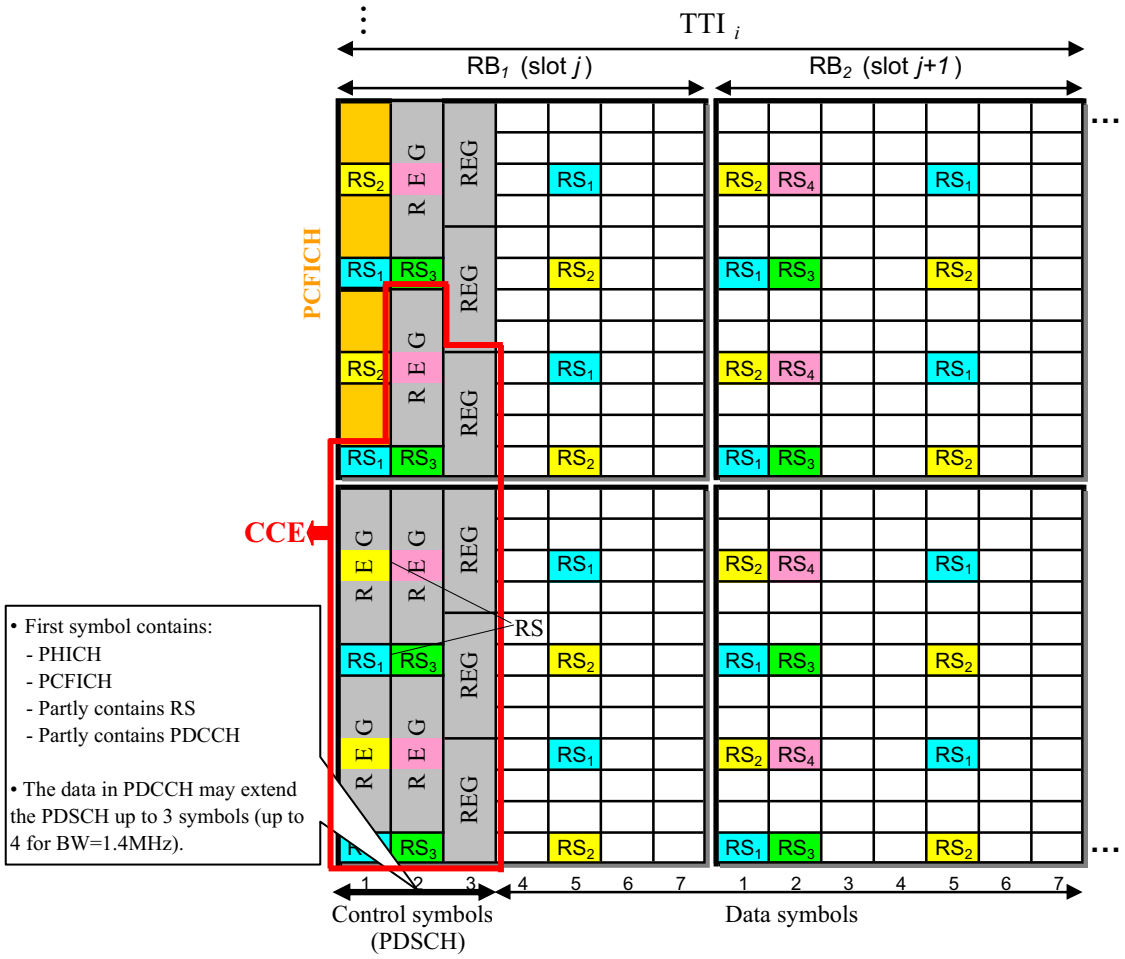


Figure 3.4: Data and Controling Symbols in Downlink TTI visualization, representing the PDCCH, PHICH, PCFICH, and RS signals, and considering the REG and CCE grouping conceptrs

in which the downlink control information is formed/packed in a standardized way, and transmitted in the PDCCH.

To accurately estimate the overhead generated by a PDCCH, $\mathcal{O}_{\text{PDCCH}}^{\text{TTI}}$ (number of REs), it is convenient to know how it is constructed. This is as follows: 1) Four REs (which have not been previously assigned to other controlling resources) are grouped forming a "Resource Element Group" (REG) [14]. 2) In turn, 9 REGs form a "Control Channel Element" (CCE). 3) Any PDCCH is transmitted on an aggregation of one or several consecutive CCEs. As a consequence, if N_{CCE} labels the number of necessary CCEs, the number of overhead REs for the PDCCH of a *single* user will be

$$\mathcal{O}_{\text{PDCCH}}^{\text{TTI, 1UE}} = 4 \cdot 9 \cdot N_{\text{CCE}} = 36 \cdot N_{\text{CCE}}. \tag{3.14}$$

The REG and CCE grouping are shown in Figure 3.4, within an example visualization of a TTI.

However, to schedule *multiple* users, a number of PDCCHs with different formats can coexist in a TTI. Let N_{UE}^{i} be the number of UEs scheduled in the i-th TTI. In the worst case, the PDCCH corresponding to each UE may have a *different* format, leading to different values of N_{CCE}^{j} per user, ($j = 1, \cdots, N_{\text{UE}}^{i}$, i being the TTI index). Thus, the number of overhead REs for the PDCCH of N_{UE}^{i} users in a generic i-th TTI will be

$$\mathcal{O}_{\text{PDCCH}}^{\text{TTI, N}_{\text{UE}}^{\text{i}}} = 36 \cdot \sum_{j=1}^{N_{\text{UE}}^{i}} N_{\text{CCE}}^{j} \tag{3.15}$$

Note in Expression (3.15) that the key point to compute accurately $\mathcal{O}_{\text{PDCCH}}^{\text{TTI, N}_{\text{UE}}^{\text{i}}}$ is the N_{CCE}^{j} value of the j-th UE.

In this respect, N_{CCE} can not take any number of REs. The LTE specification [159] states that N_{CCE} can only be either 1, 2, 4, or 8 REs, depending on the required N_{REG}. In turn, N_{REG} is related to the "PDCCH bit length", L_{PDCCH}, which is nothing but the necessary number of bits to transport the DCI corresponding to each UE. The specification [159] defines four different intervals for L_{PDCCH}, which corresponds to four PDCCH "formats". Table 3.6 shows the N_{CCE} allowed sizes (1, 2, 4, or 8 REs), and the corresponding PDCCH formats ("0", "1", "2", or "3") as a function of the allowed L_{PDCCH} intervals.

The question arising here is how L_{PDCCH} can be computed. Since a PDCCH is made of the concatenation of one DCI with 16 scrambled CRC bits, followed by a coding with rate 1/3, then

$$L_{\text{PDCCH}} = 3 \cdot (L_{\text{DCI}} + 16) \quad \text{(bit)} \tag{3.16}$$

where L_{DCI} is the DCI bit length, or number of bits that are necessary to encode the DCI information.

Table 3.6: PDCCH Construction Parameters [159]: N_{CCE}, N_{REG}, and PDCCH Format as a function of the L_{PDCCH} interval. See [14] for further details.

L_{PDCCH} interval	N_{CCE}	N_{REG}	PDCCH Format
up to 72	1	9	0
$72 < L_{\mathrm{PDCCH}} \leqslant 144$	2	18	1
$144 < L_{\mathrm{PDCCH}} \leqslant 288$	4	36	2
$288 < L_{\mathrm{PDCCH}} \leqslant 576$	8	72	3

The exact bit level content of each DCI format is defined in Section 5.3.3 of the 3GPP standard [148]. The DCI length L_{DCI}[bits] is a function of the system BW (via N_{RB}), and of the DCI formats f_{DCI} (summarized in Table 3.5). Being more concrete, Table 3.7 lists the expression to compute L_{DCI}. It depends on: 1) N_{RB} and P (or equivalently, on the BW –Table 3.3–); 2) The number of bits (a, b, and c) used for indicating the number of transmission antennae to be use (TPMI precoding information), and whose values have been listed in Table 3.8; and 3) The number of transmitted codewords (via parameter q), which is a function of the resource scheduling and the multi-antenna transmission scheme (Table 3.5).

Table 3.7: Expressions to compute the DCI length (L_{DCI} [bits]) as a function of the downlink scheduling type [14]. The values of a, b, and c, used for indicating the number of transmission antennae are listed in Table 3.8. Table 3.5 describes the DCI formats.

Scheduling type	DCI format	L_{DCI}(bit)
Type 0	1	$\lceil N_{\mathrm{RB}}/P \rceil + 16$
	2	$\lceil N_{\mathrm{RB}}/P \rceil + 7 + a + 8 \cdot (q+1)$
	2A	$\lceil N_{\mathrm{RB}}/P \rceil + 7 + b + 8 \cdot (q+1)$
Type 1	1	$\lceil N_{\mathrm{RB}}/P \rceil + 16$
	2	$\lceil N_{\mathrm{RB}}/P \rceil + 7 + a + 8 \cdot (q+1)$
	2A	$\lceil N_{\mathrm{RB}}/P \rceil + 7 + b + 8 \cdot (q+1)$
Type 2	1	$\lceil \log_2(N_{\mathrm{RB}} \cdot (N_{\mathrm{RB}}+1)/2) \rceil + 13$
	2	$\lceil \log_2(N_{\mathrm{RB}} \cdot (N_{\mathrm{RB}}+1)/2) \rceil + 15 + c$
	2	$\lceil \log_2(N_{\mathrm{RB}} \cdot (N_{\mathrm{RB}}+1)/2) \rceil + 16$
	2A	$\lceil \log_2(N_{\mathrm{RB}} \cdot (N_{\mathrm{RB}}+1)/2) \rceil + 16 + c$

Table 3.8: Values of parameters a, b, and c to be used in Table 3.7.

MIMO scheme	a	b	c
2-TX	3	0	2
4-TX	6	2	4

For the sake of clarity, Figure 3.5 summarizes the steps to compute $\mathcal{O}_{\mathrm{PDCCH}}^{\mathrm{TTI,\ N_{UE}^{i}}}$.

Modeling the Reference Signal Overhead

Reference Signals (RSs) are used for transmitting information related to the required estimations of channel quality and signal level. As explained in [14], the number of RSs per TTI depends on the number of Tx antennas to be used:

$$N_{\mathrm{RS|(1\ TTI)}} = \begin{cases} 8, \text{ for } n = 1 \\ 16, \text{ for } n = 2 \\ 24, \text{ for } n = 4 \end{cases} \tag{3.17}$$

If the user is allowed to use N_{RB} in a given TTI, then the number of REs *per TTI* reserved for RS-overhead will be

$$\mathcal{O}_{\mathrm{RS}}^{\mathrm{TTI}} = N_{\mathrm{RS|(1\ TTI)}}(n) \cdot N_{\mathrm{RB}} \tag{3.18}$$

Apart from these cell-specific RSs (often referred to as "common" RSs, as they are available to all UEs in a cell [14]), which are present in all data transmitting DL subframes, there are other reference signals that only appear in specific conditions: 1) UE-specific RSs, which may be embedded in the data for specific UEs; and 2) Multimedia Broadcast Single Frequency Network (MBSFN)-specific RSs, which are only used for Multimedia Broadcast Single Frequency Network (MBSFN) operation, to transmit multicast or broadcast data as a multi-cell transmission over a synchronized Single Frequency Network (SFN) [14]. For improving the performance of SFN, ultra-extended cyclic prefix duration has been recently considered [161] to account for the larger delay spread, especially in rural and hilly areas. This specific multi-cell transmission over SFN, using ultra-extended cyclic prefix, is not considered in general overhead studies [14].

Modeling the Physical Hybrid ARQ Indicator Channel Overhead

Physical Hybrid ARQ Indicator Channel (PHICH) carries the Hybrid-ARQ acknowledgment in DL (corresponding to its previous counterpart uplink data transfers). The number of overhead REs for PHICH can be computed as follows.

A PHICH is placed in the first symbol of each TTI, and is always transmitted using BPSK. Each PHICH carries in turn 3 ACK bits (111) or NACK (000), which require 3 REs each (with BPSK modulation). In turn, in each TTI, up to 8 PHICHs can be grouped forming one "PHICH group". The number of PHICH groups ($N_{\mathrm{PHICH}}^{\mathrm{groups}}$) is a function of the system BW (via N_{RB}) and is fixed in all subframes. The 3GGP standard has stated an expression to compute the maximum value for $N_{\mathrm{PHICH}}^{\mathrm{groups}}$, which for normal CP, is [159, 162]

$$N_{\mathrm{PHICH}}^{\mathrm{groups}} = \left\lceil N_g \cdot \left(\frac{N_{\mathrm{RB}}}{8}\right) \right\rceil \tag{3.19}$$

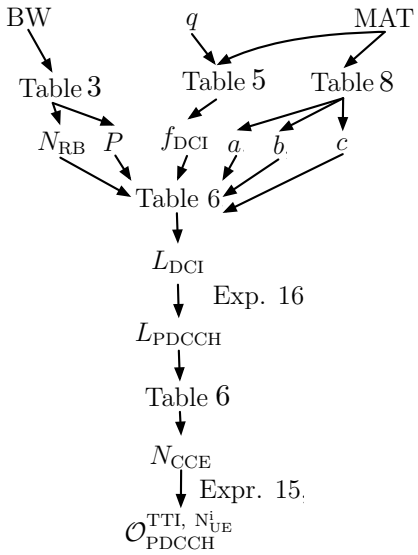


Figure 3.5: Summary of steps to compute $\mathcal{O}_{\text{PDCCH}}^{\text{TTI, N}_{\text{UE}}^{\text{i}}}$.

where N_g is the "PHICH Group Scaling Factor" [162], and can only take values in the set $\{1/6, 1/2, 1, 2\}$, as shown in Table 3.9. It lists the number of PHICH groups as a function of BW and N_g.

Table 3.9: Number of PHICH groups as a function of BW and N_g. Information obtained from 3GPP TS 36.211 [162].

Number of PHICH groups, $N_{\text{PHICH}}^{\text{groups}}$						
	BW (MHz)					
Ng	1.4	3	5	10	15	20
1/6	1	1	1	2	2	3
1/2	1	1	2	4	5	7
1	1	2	4	7	10	13
2	2	4	7	13	19	25

Since 1 PHICH group requires 12 REs –because the 8 PHICHs forming a PHICH group differ from each other by orthogonal sequence indexes, which impose a spreading factor of 4 over the 3-bit PHICH– then the overhead caused by PHICH per TTI and codeword is simply

$$\mathcal{O}_{\text{PHICH}}^{\text{TTI, codeword}} = N_{\text{PHICH}}^{\text{groups}} \cdot 12 \quad \text{(REs)} \tag{3.20}$$

We use uperscript "codeword" in Expression (3.20) to emphasize that the HARQ process is carried out separately for each codeword. To include the number of codewords, Expression (3.20) can be modified as follows:

$$\mathcal{O}_{\text{PHICH}}^{\text{TTI}} = (q+1) \cdot N_{\text{PHICH}}^{\text{groups}} \cdot 12 \quad \text{(REs)} \tag{3.21}$$

Modeling the Physical Format Indicator Channel Overhead

The DL Physical Control Format Indicator Channel (PCFICH) is one of the control channels that works at physical level. It is used to dynamically indicate the number of OFDM symbols to be used by the PDCCH to transmit the *controlling* information in a subframe.

Since the PCFICH is transmitted in the first symbol of each subframe (using QPSK modulation), and considering that it contains up to 16 REs per TTI, then the corresponding overhead is

$$\mathcal{O}_{\text{PCFICH}}^{\text{TTI}} = 16 \text{ REs}. \tag{3.22}$$

3.4.2 Overhead mechanisms with period 1 frame

Modeling the Synchronization Signal Overhead

Cell synchronization is a procedure for an UE to acquire the necessary information when connecting to an eNB. This synchronization process, which involves two synchronization channels (SCH) –Primary Synchronization channel (P-SCH) and Secondary Synchronization channel (S-SCH)–, consists basically of two steps [14]. In the first one, the UE searches for a signal called "Primary Synchronization Signal" (PSS) [163–165], which enables the UE to be synchronized on subframe level. In a second step, the UE uses the Secondary Synchronization Signal (SSS) to obtain the PHY layer cell identity group number. The following details [14] are useful to compute the synchronization overhead. On one hand, transmitting PSS requires to reserve a number of REs: Part of the PSS information is transmitted on 2 slots (in the last symbol of slots 0 and 10) of each frame, over 62 subcarriers centered in the Direct Current (DC) SC (a subcarrier that does not transport information, and is used by the UE to locate the center of the OFDM frequency band). Thus, it is necessary to reserve 62 REs per slot. Additionally, another 5 REs before and after the P-SCH allocation (in each symbol on the 62 subcarriers) are also reserved: $(5+5)$ REs/slot. As a conclusion: PSS overhead needs 2 slots/frame $\times$ $(5 + 62 + 5)$ REs/slot = 144 REs per frame. On the other hand, but in a similar line of reasoning, SSS requires to reserve 2 slots/frame $\times$ $(5 + 62 + 5)$ REs/slot = 144 REs per frame. Thus, the total overhead cause by to both PSS and SSS is

$$\mathcal{O}_{\text{SCH}}^{\text{frame}} = 288 \text{ (REs per frame)} \tag{3.23}$$

Note that $\mathcal{O}_{\text{SCH}}^{frame}$ is used (repeated) only once per frame, unlike the previous one, repeated once per TTI.

Modelling the Broadcast Channel Overhead

The Physical Broadcast Channel (PBCH) is transmitted over 4 symbols (from 0 to 3) of the slot 1 of each frame, on 72 subcarriers centered in the DC subcarrier. Thus, the corresponding number of REs that must be reserved for PBCH in a frame is: 4 symbols/slot $\times$ 1 slot/frame $\times$ 72 RE/frame = 288 REs per frame. That is, PBCH overhead is:

$$\mathcal{O}_{\text{PBCH}}^{\text{frame}} = 288 \text{ (REs per frame)} \tag{3.24}$$

3.5 Total overhead $\mathcal{O}_{\text{TOT}}^{\Delta t}$

At this point, and once we have modeled all its components, superscript Δt used in Expression (3.2) for labeling $\mathcal{O}_{\text{TOT}}^{\Delta t}$ acquires a new clearer meaning. This is because the global (total) overhead $\mathcal{O}_{\text{TOT}}^{\Delta t}$ depends on the length of Δt since there are some REs (see for instance, $\mathcal{O}_{\text{SCH}}^{frame}$ in (3.23) and $\mathcal{O}_{\text{PBCH}}^{frame}$ in (3.24)) that do *not* appear in all the TTIs but only at a longer scale, that is, "at frame scale". Transmitting a long number of user bytes (a big download) that would require to use several frames should demand to compute the overhead at frame scale.

On one hand, the total number of REs in one TTI, $\mathcal{O}_{\text{RE}}^{\Delta t=\text{TTI}}$, is simply the sum of all its components:

$$\mathcal{O}_{\text{TOTAL}}^{\text{TTI}} = \mathcal{O}_{\text{PCFICH}}^{\text{TTI}} + \mathcal{O}_{\text{PHICH}}^{\text{TTI}} + \mathcal{O}_{\text{PDCCH}}^{\text{TTI}} + \mathcal{O}_{\text{RS}}^{\text{TTI}} \tag{3.25}$$

On the other hand, when working at frame scale, as shown in Subsections 3.4.2 and 3.4.2, there are other controlling signals (such as PBCH and SCH) that must be added since they appear once per frame. These overheads are neither repeated in any generic TTI nor included in the initial overhead symbols of each TTI. Computing the overhead throughout a generic frame, $\mathcal{O}_{\text{RE}}^{\Delta t=frame}$, requires to sum all the overhead contributions corresponding to the ten TTIs that compose a frame. Note that a each TTI may contain different scheduling configuration and, as a consequence, the number of overhead REs of each TTI, may be different from each other. If i indexes the TTIs (with $i = 0, 1, \cdots, 9$), then the number of REs in a complete frame will be

$$\mathcal{O}_{\text{TOT}}^{frame} = \mathcal{O}_{\text{SCH}}^{frame} + \mathcal{O}_{\text{PBCH}}^{frame} + \sum_{i=0}^{9} \mathcal{O}_{\text{TOT}}^{\text{TTI}\ i} \tag{3.26}$$

3.6 Overhead and Initial Controlling Symbols

Since the three classes of controlling symbols defined by LTE standards for the Physical DL Shared Channel (PDSCH) *always* start at the *first* symbol of each TTI, they will be called "Initial Controlling Symbols" (ICS) henceforth. As was illustrated in Figure 3.4, the ICS covers PDCCH, PHICH, PCFICH and and a fraction α of RS signals. For reasons that will appear clearer some paragraphs below, we formally group the overhead components corresponding to the ICS during a TTI as

$$\mathcal{O}^{\text{TTI}}_{\text{PCFICH}} + \mathcal{O}^{\text{TTI}}_{\text{PHICH}} + \mathcal{O}^{\text{TTI}}_{\text{PDCCH}} + \alpha \cdot \mathcal{O}^{\text{TTI}}_{\text{RS}} \equiv \mathcal{O}^{\text{TTI}}_{\text{ICS}} \text{ (RE)}, \tag{3.27}$$

where $\mathcal{O}^{\text{TTI}}_{\text{ICS}}$ labels the number of REs used for ICS within a TTI, and factor α represents the fraction of the RSs that are included in the ICS for each TTI. α depends on the number of transmitting antennas n as

$$\alpha(n) = \begin{cases} 1/4, & \text{for} \quad n = 1, 2 \\ 1/3, & \text{for} \quad n = 4 \end{cases} \tag{3.28}$$

Furthermore, $\mathcal{O}^{\text{TTI}}_{\text{ICS}}$ formula in Expression (3.27) can be rewritten to consider explicitly the influence of N_{CCE} since, in the worst case, any UE may have different N^j_{CCE}, ($j = 1, \cdots, N^i_{\text{UE}}$, i being the TTI index). Taking into account that $\mathcal{O}^{\text{TTI}}_{\text{PDCCH}}$ is given by Exp. (3.15), then Expression (3.27) can be reformulated as

$$\begin{aligned} \mathcal{O}^{\text{TTI}}_{\text{ICS}} &\equiv \left(\mathcal{O}^{\text{TTI}}_{\text{PCFICH}} + \mathcal{O}^{\text{TTI}}_{\text{PHICH}} + \alpha \cdot \mathcal{O}^{\text{TTI}}_{\text{RS}}\right) + \mathcal{O}^{\text{TTI}}_{\text{PDCCH}} \equiv \\ &\equiv \Gamma + \mathcal{O}^{\text{TTI, N}^{\text{i}}_{\text{UE}}}_{\text{PDCCH}} = \Gamma + 36 \cdot \sum_{j=1}^{N^i_{\text{UE}}} N^j_{\text{CCE}}, \end{aligned} \tag{3.29}$$

Γ being

$$\Gamma = \mathcal{O}^{\text{TTI}}_{\text{PCFICH}} + \mathcal{O}^{\text{TTI}}_{\text{PHICH}} + \alpha \cdot \mathcal{O}^{\text{TTI}}_{\text{RS}}, \tag{3.30}$$

which do *not* depend on the number of users for CRC bits calculation.

Although not explicitly expressed in Expression (3.27), please note that: 1) $\mathcal{O}^{\text{TTI}}_{\text{PHICH}}$ is a function of N_{RB} and N^g_{PHICH}; 2) $\mathcal{O}^{\text{TTI}}_{\text{PDCCH}}$ depends on N_{RB}, f_{DCI}, n-TX, and N_{UE}; and 3) $\mathcal{O}^{\text{TTI}}_{\text{RS}}$ and α are functions of n-TX.

The minimum number of controlling *symbols* that can accommodate the above calculated controlling REs in a generic TTI will be simply

$$\mathcal{O}^{\text{TTI}}_{\text{ICS, symb}} = \left\lceil \frac{\mathcal{O}^{\text{TTI}}_{\text{ICS}}}{12 \cdot N_{\text{RB}}} \right\rceil \text{ (symbols)} \tag{3.31}$$

because in a RB there are $N^{\text{RB}}_{\text{SC}} = 12$ subcarriers.

From a practical standpoint, its importance lies in the fact that the standard defines a maximum number of *controlling symbols* for the Physical DL Shared Channel

(PDSCH) per subframe, (labelled ξ in this paper), as a function of the bandwidth used. The standard states:

$$\xi = \begin{cases} 4, & \text{for BW} = 1.4 \quad \text{MHz} \\ 3, & \text{otherwise} \end{cases} \tag{3.32}$$

so that the rest of the symbols in the subframe can be used for data transmission.

Thus, it is necessary for $\mathcal{O}^{\mathrm{TTI}}_{\mathrm{ICS,\ symb}}$ to fulfill

$$\mathcal{O}^{\mathrm{TTI}}_{\mathrm{ICS,\ symb}} \leqslant \xi, \tag{3.33}$$

This constrain has two practical consequences. The first one is that it limits the number of UEs that can be scheduled within a TTI (Section 3.7). The second consequence can lead to LTE configurations in which there could be a number of wasted resources in the initial controlling symbols when the symbols are not fully filled up with the controlling REs. In these cases, computing accurately the overhead and the capacity demand a methodology that we described in Section 3.8.

3.7 Overhead and maximum number of UEs (with identical N_{CCE} value) per TTI

Basically, the existence of a maximum number of UEs per TTI lies on the fact that there is a maximum number of controlling OFDM symbols for shared-data transmitting subframes (up to $\xi = 4$ for BW of 1.4MHz) (see Expression (3.32)). This figure is different from the maximum number of UEs supported by a cell, as the active users can be scheduled in distributed TTIs.

The $\mathcal{O}^{\mathrm{TTI}}_{\mathrm{ICS,\ symb}}$ calculation in Exp. (3.31) reveals a practical restriction on the number of UEs that can be scheduled within a specific system configuration. That is, if the computed $\mathcal{O}^{\mathrm{TTI}}_{\mathrm{ICS}}$ for a TTI is greater than $\xi \cdot 12 \cdot N_{\mathrm{RB}}$ (where $\xi = 4$ for 1.4 MHz and $\xi = 3$ otherwise) the imposed initial controlling symbols will exceed the maximum permitted, and as a consequence, the number of scheduled UEs in that TTI would have to be decreased.

To gain some insight, we can explore the following limiting case in which all the N^i_{UE} UEs are assumed to be *identical*, thus having identical PDCCH format and identical N_{UE}, then:

1. Expression (3.15) become into $\mathcal{O}^{\mathrm{TTI,\ N^i_{UE}}}_{\mathrm{PDCCH}} = 36 \cdot N_{\mathrm{CCE}} \cdot N^i_{\mathrm{UE}}$. By substituting it in Expression (3.29), then

$$\mathcal{O}^{\mathrm{TTI}}_{\mathrm{ICS}} = \Gamma + 36 \cdot N_{\mathrm{CCE}} \cdot N^i_{\mathrm{UE}} \tag{3.34}$$

2. The constrain (3.33) fulfills as:

$$\mathcal{O}_{\mathrm{ICS}}^{\mathrm{TTI}} = \xi \cdot 12 \cdot N_{\mathrm{RB}} \tag{3.35}$$

From Equations (3.34) and (3.35), we obtain the maximum number of *identical* users (with the same type of service and N_{CCE}):

$$N_{\mathrm{UE}}^{\mathrm{TTI}_i}|_{\mathrm{MAX}} = \frac{\lceil \xi \cdot 12 \cdot N_{\mathrm{RB}} - \Gamma \rceil}{36 \cdot N_{\mathrm{CCE}}} \tag{3.36}$$

3.8 Realistic scenario: multiple users with different services

Unlike the previous section, it is likely that most users have different receiving characteristics and services. This makes more difficult to apply the aforementioned model since it is based on some simplifying assumptions that we remove here when presenting the methodology that follows.

The proposed methodology is based on describing the overhead over a TTI $\mathcal{O}_{\mathrm{TOT}}^{\mathrm{TTI}}$ *on the basis on the number of initial controlling symbols*, which, as mentioned before, are constrained by inequality (3.33). Apart from the initial controlling symbols described in Section 3.6, there are a number of REs (used for reference signals), $(1-\alpha)\cdot\mathcal{O}_{\mathrm{RS}}^{\mathrm{TTI}}$, which have not been reserved for controlling symbols and that must be account for. This is why $\mathcal{O}_{\mathrm{TOT}}^{\mathrm{TTI}}$ can alternatively be written as

$$\mathcal{O}_{\mathrm{TOT}}^{\mathrm{TTI}} = 12 \cdot N_{\mathrm{RB}} \cdot \mathcal{O}_{\mathrm{ICS,\ symb}}^{\mathrm{TTI}} + (1-\alpha)\cdot\mathcal{O}_{\mathrm{RS}}^{\mathrm{TTI}}, \tag{3.37}$$

which, using Expression (3.31) for $\mathcal{O}_{\mathrm{ICS,\ symb}}^{\mathrm{TTI}}$, leads to the total overhead stated by (3.25). However, from an operative viewpoint, Exp. (3.37) will allow us to proceed further in complex scenarios as follows.

Let us assume that, in the $i-th$ TTI, there are $N_{\mathrm{UE}}^{\mathrm{TTI}_i}$ users to be scheduled. Let $N_{\mathrm{CCE},j}^{\mathrm{TTI}_i}$ be the N_{CCE} value of UE number j, with $j = 1, \cdots, N_{\mathrm{UE}}^{\mathrm{TTI}_i}$. Note that each $N_{\mathrm{CCE},j}^{\mathrm{TTI}_i}$ could be different from each other because, in general, the users can have different characteristics (CQI, services, etc.). The methodology is as follows:

1. Compute $\mathcal{O}_{\mathrm{PDCCH}}^{\mathrm{TTI,\ N_{UE}^{TTI_i}}}$ using Expression (3.15)
2. Calculate the number of overhead REs for ICS, $\mathcal{O}_{\mathrm{ICS}}^{\mathrm{TTI}}$, using Formula (3.29).
3. Compute the maximum integer that fulfills (3.33). Let $(\mathcal{O}_{\mathrm{ICS,\ symb}}^{\mathrm{TTI}})^{\mathcal{M}}$ be that integer.
4. Substitute $(\mathcal{O}_{\mathrm{ICS,\ symb}}^{\mathrm{TTI}})^{\mathcal{M}}$ into Expression (3.37), leading to

$$(\mathcal{O}_{\mathrm{TOT}}^{\mathrm{TTI}})^{\mathcal{M}} = 12 \cdot N_{\mathrm{RB}} \cdot (\mathcal{O}_{\mathrm{ICS,\ symb}}^{\mathrm{TTI}})^{\mathcal{M}} + (1-\alpha)\cdot\mathcal{O}_{\mathrm{RS}}^{\mathrm{TTI}}, \tag{3.38}$$

which allows for estimating the maximum number of useful REs,

$$(\mathcal{U}_{\mathrm{RE}}^{\mathrm{TTI}_i})^{\mathcal{M}} = \mathcal{A}_{\mathrm{RE}} - (\mathcal{O}_{\mathrm{TOT}}^{\mathrm{TTI}})^{\mathcal{M}} \tag{3.39}$$

5. Introduce $(\mathcal{U}_{\mathrm{RE}})^{\mathcal{M}}$ into Expression (3.9), obtaining

$$\begin{aligned} \mathrm{TBS}_i = \sum_{j=1}^{N_{\mathrm{UE}}^{\mathrm{TTI}_i}} \Omega_j \cdot \left(w_j \cdot (\mathcal{U}_{\mathrm{RE}}^{\mathrm{TTI}_i})^{\mathcal{M}} \cdot Q_j \cdot \mathrm{ECR}_j^{\mathrm{MCS}} - 24 \cdot N_{\mathrm{UE}}^{\mathrm{TTI}_i} \right) \\ \equiv \sum_{j=1}^{N_{\mathrm{UE}}^{\mathrm{TTI}_i}} (\mathrm{TBS})_j^{\mathcal{M}} \end{aligned} \tag{3.40}$$

where $(\mathrm{TBS})_j^{\mathcal{M}}$ is

$$(\mathrm{TBS})_j^{\mathcal{M}} \cdot \Omega_j \doteq \left(w_j \cdot (\mathcal{U}_{\mathrm{RE}}^{\mathrm{TTI}_i})^{\mathcal{M}} \cdot Q_j \cdot \mathrm{ECR}_j^{\mathrm{MCS}} - 24 \cdot N_{\mathrm{UE}}^{\mathrm{TTI}_i} \right) \tag{3.41}$$

6. Using Expression (3.1), compute the corresponding throughput for any user j as:

$$(\mathrm{TBS})_j^{\mathcal{M}} \cdot 10^{-3} \doteq \mathcal{T}_j \text{ (Mbps)} \tag{3.42}$$

3.9 Experimental work

We start our experimental work by considering, as a reference, the "best" case that will be described in Subsection 3.9.1. It has been obtained by using the proposed methodology with the optimum parameters that *minimize* the total overhead. The reason why we present this case first is to establish a framework that aims to compare it to other more realistic cases, in which we will progressively remove the optimum parameters by others that could be used during the normal operation of an LTE network (Subsection 3.9.2). This will assist us in providing more accurate estimation of overhead under real operation conditions, and in computing to what extent this affects the peak data rate, the maximum number of users per TTI (Subsection 3.9.4) or users' throughput in a multi service scenario (Subsection 3.9.5).

3.9.1 Reference case

Considering that one single UE is using all resources in each Δt, the Rference Case, corresponds to a 4×4 MIMO configuration ($n = 4$) with the values of the involved parameters listed in Table 3.10. These are those that minimize the total overhead $\mathcal{O}_{\mathrm{TOT}}^{\Delta t}$ (RE).

Table 3.10: Values of the parameters that minimize the total overhead in the reference study case. See Table 3.5 for further details.

Scheduling type	DCI format f_{DCI}	q	Ω	CQI	N_g	CQI	ECR_{MCS}
0 (RGB based)	2A (Open Loop)	1	4	15	1/6	15	0.926

Figure 3.6 shows the total number of overhead REs required within a time interval Δt, $\mathcal{O}_{\text{TOT}}^{\Delta t}$, either 1 TTI or 1 frame, as a function of the scalable LTE BW (MHz). Note that the overhead at *frame scale* (blue $\square$ symbols), $\mathcal{O}_{\text{TOT}}^{\text{frame}}$, is higher than the overhead computed at TTI scale, $\mathcal{O}_{\text{TOT}}^{\text{TTI}}$ (red $\bigcirc$ symbols). This makes sense since a frame contains overhead components occurring once per frame ($\mathcal{O}_{\text{PBSH}}^{\text{frame}}$ and $\mathcal{O}_{\text{SCH}}^{\text{frame}}$) which do *not* exist in a generic TTI. On the other hand, and as expected, both $\mathcal{O}_{\text{TOT}}^{\text{frame}}$ and $\mathcal{O}_{\text{TOT}}^{\text{TTI}}$ increase as the BW to be used increases. However, this *absolute* measure could be misleading. The reason is that, as the BW increases, the number of total available REs $\mathcal{A}_{\text{TOT}}^{\Delta t}$ also increases, and so does $\mathcal{O}_{\text{TOT}}^{\Delta t}$. A *relative* measure will assist us in better illustrating such an influence. This is why we use the percentage of overhead relative to total available resources

$$\Delta\mathcal{O}_{\text{TOT}}^{\Delta t}(\%) = \frac{\mathcal{O}_{\text{TOT}}^{\Delta t}}{\mathcal{A}_{\text{TOT}}^{\Delta t}} \times 100. \tag{3.43}$$

Regarding this, Figure 3.7 shows metric $\Delta\mathcal{O}_{\text{TOT}}^{\Delta t}$ as a function of the channel bandwidth, BW(MHz), and parametrized by the interval time (either 1 TTI of 1 frame) during which the overhead is compute, Δt. Blue $\square$ symbols represent $\Delta\mathcal{O}_{\text{TOT}}^{\text{frame}}$, while red $\bigcirc$ symbols correspond to $\Delta\mathcal{O}_{\text{TOT}}^{\text{TTI}}$. For comparative purposes we have also represented two overhead results found in the review: 14.3% [72] (black $\Diamond$ symbol for BW = 20 MHz) and 12.12% [66] ($\bigstar$ symbol for BW = 10 MHz).

Contrary to what Figure 3.6 suggests, Figure 3.7 reveals that:

1. The relative influence of overhead *reduces* as BW increases. Note that for the maximum bandwidth value, BW = 20 MHz (the bandwidth for which LTE is able to reach its maximum bit data rates), the percentage (with respect to the total available resources) of both computed overhead, $\Delta\mathcal{O}_{\text{TOT}}^{\text{frame}}$ and $\Delta\mathcal{O}_{\text{TOT}}^{\text{TTI}}$ tend to an *upper bound overhead* $\approx 15.9\%$.

2. While, for BW= 1.4MHz, $\Delta\mathcal{O}_{\text{TOT}}^{\text{frame}}\,|_{(\text{BW=1.4MHz})} = 31.1\%$ is considerable higher than $\Delta\mathcal{O}_{\text{TOT}}^{\text{TTI}}\,|_{(\text{BW=1.4MHz})} = 25.4\%$, however, as BW increases their respective values tend to become closer so that, for the maximum BW value (BW = 20MHz), $\Delta\mathcal{O}_{\text{TOT}}^{\text{frame}}\,|_{(\text{BW=20MHz})} \approx 15.9\%$ and $\Delta\mathcal{O}_{\text{TOT}}^{\text{TTI}}\,|_{(\text{BW=20MHz})} \approx 15.8\%$.

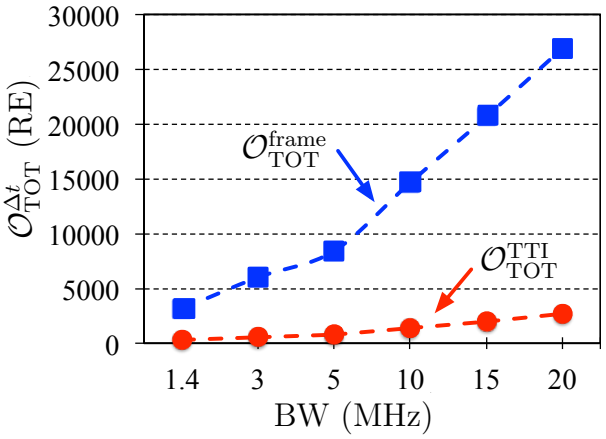


Figure 3.6: Total overhead $\mathcal{O}_{\mathrm{TOT}}^{\Delta t}$ (RE) as a function of BW (MHz). Blue □-symbols represent $\mathcal{O}_{\mathrm{TOT}}^{\mathrm{frame}}$, while red ◯-symbols show the overhead at TTI scale, $\mathcal{O}_{\mathrm{TOT}}^{\mathrm{TTI}}$.

3. The results found in research works [66, 72] *underestimate* overhead because because they refer *only* to RS contributions: 12.12% (★ symbol for BW = 10 MHz corresponding to [66]) compared to our (total) result $\Delta\mathcal{O}_{\mathrm{TOT}}^{\mathrm{frame}}\,|_{(\mathrm{BW=10MHz})}$= 15.9%, and 14.3% (black ◊ symbol for BW = 20 MHz estimated in [72]) compared to our result $\Delta\mathcal{O}_{\mathrm{TOT}}^{\mathrm{frame}}\,|_{(\mathrm{BW=20MHz})}$= 15.6%.

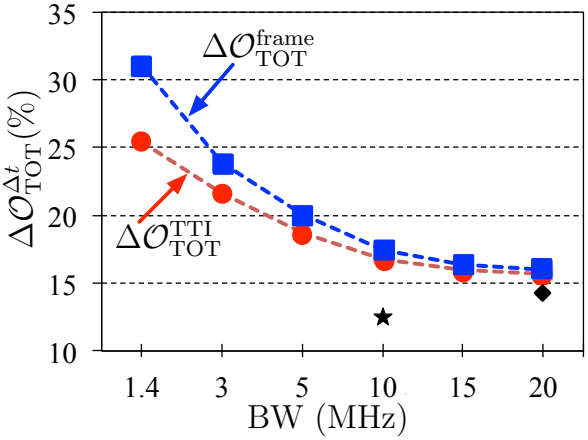


Figure 3.7: $\Delta\mathcal{O}_{\mathrm{TOT}}^{\Delta t}$(%) as a function of the bandwidth BW (MHz). Blue square symbols (□) represent $\Delta\mathcal{O}_{\mathrm{TOT}}^{\mathrm{frame}}$, while red circular symbols (◯) represent $\Delta\mathcal{O}_{\mathrm{TOT}}^{\mathrm{TTI}}$. Black ◊ symbol and ★ correspond to those overhead estimated in [72] and [66], respectively.

As done in some works, whether or not some components can be disregarded

is a topic of practical interest because it would simplify the model. Thus the next question arising is to what extent each overhead components have different impacts on the total overhead, or, in other words, which is the contribution of its components. To quantify their relative weights on the total overhead (both on $\mathcal{O}_{\text{TOT}}^{\text{TTI}}$ and on $\mathcal{O}_{\text{TOT}}^{\text{frame}}$), we have defined

$$\Delta\mathcal{O}_{\text{comp }i}^{\text{TTI}}(\%) = \frac{\mathcal{O}_{\text{comp }i}^{\text{TTI}}}{\mathcal{O}_{\text{TOT}}^{\text{TTI}}} \times 100 \tag{3.44}$$

$$\Delta\mathcal{O}_{\text{comp }i}^{\text{frame}}(\%) = \frac{\mathcal{O}_{\text{comp }i}^{\text{frame}}}{\mathcal{O}_{\text{TOT}}^{\text{frame}}} \times 100 \tag{3.45}$$

where subscript "comp i" refers to any of the overhead components (RS, PDCCH, etc.), either at TTI level (superscript "TTI" in Exp. (3.44)) or at frame level (superscript "frame" in (3.45)).

Using metrics (3.44) and (3.45), Figures 3.8 (a) and (b) show the relative contribution (with respect to the total overhead) of each component computed, respectively, at TTI scale and at frame scale.

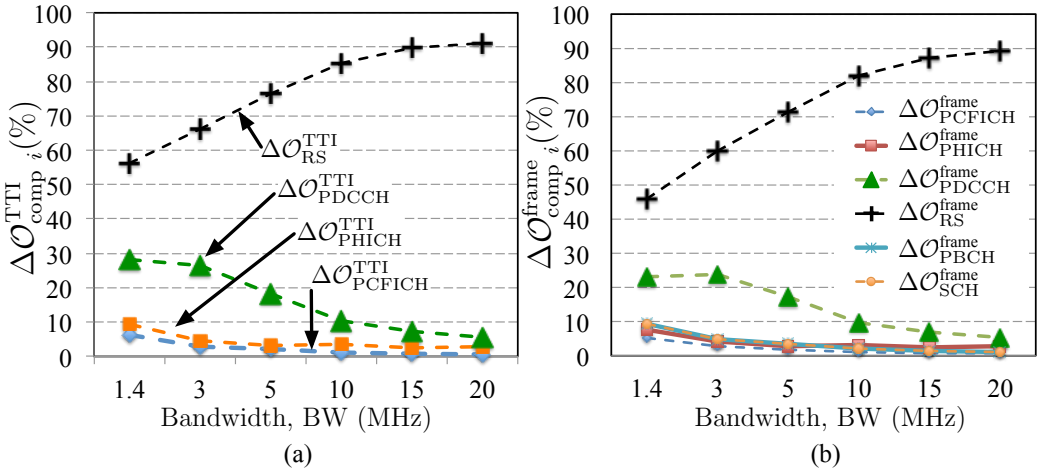


Figure 3.8: (a) Percentage of each overhead component $\Delta\mathcal{O}_{\text{comp }i}^{\text{TTI}}(\%)$ with respect to the total overhead at TTI scale $\mathcal{O}_{\text{TOTAL}}^{\text{TTI}}$, as a function of BW (MHz). (b) Percentage of each overhead component $\Delta\mathcal{O}_{\text{comp }i}^{\text{frame}}(\%)$ with respect to the total overhead at frame scale $\mathcal{O}_{\text{TOTAL}}^{\text{frame}}$, as a function of BW (MHz). The inset defines the symbols used for any component.

Specifically, Figure 3.8 (a), which shows the percentage of each component over the total overhead $\mathcal{O}_{\text{TOT}}^{\text{TTI}}$ (using the metric defined in (3.44)), reveals that $\mathcal{O}_{\text{RS}}^{\text{TTI}}$ plays the key role since is much higher than all the other components in all the scalable LTE bandwidths. Furthermore, note that, as BW increases, $\mathcal{O}_{\text{RS}}^{\text{TTI}}$ increases, while the others reduce. For BW = 20MHz, this element is as large as $\Delta\mathcal{O}_{\text{RS}}^{\text{TTI}} = 91\%$.

Furthermore, $\Delta\mathcal{O}_{\mathrm{RS}}^{\mathrm{TTI}}+\Delta\mathcal{O}_{\mathrm{PDCCH}}^{\mathrm{TTI}}$ ranges between 84% and 97% in all the BW range 1.4MHz $\leq$ BW $\leq$ 20MHz. Figure 3.8 (b), computed at frame scale, exhibits the same trends, although the relative contributions of $\mathcal{O}_{\mathrm{RS}}^{\mathrm{frame}}$ and $\mathcal{O}_{\mathrm{PDCCH}}^{\mathrm{frame}}$ are slightly smaller. Regarding this, note that: 1) for BW = 20MHz, $\Delta\mathcal{O}_{\mathrm{RS}}^{\mathrm{frame}} = 89.23\%$; 2) $\Delta\mathcal{O}_{\mathrm{RS}}^{\mathrm{frame}}+\Delta\mathcal{O}_{\mathrm{PDCCH}}^{\mathrm{frame}}$ ranges between 68.66% and 94.58% $\forall$ BW $\in$ [1.4MHz, 20MHz]. This is because, at frame scale, there are overhead components ($\mathcal{O}_{\mathrm{SCH}}^{\mathrm{frame}}$ and $\mathcal{O}_{\mathrm{PBCH}}^{\mathrm{frame}}$) that do *not* appear in a generic TTI, reducing thus the relative contribution of $\mathcal{O}_{\mathrm{RS}}^{\mathrm{frame}}$ and $\mathcal{O}_{\mathrm{PDCCH}}^{\mathrm{frame}}$, which, in any case, are clearly the dominant contributions to the total overhead.

Finally, the absolute, total overheads ($\mathcal{O}_{\mathrm{TOT}}^{\mathrm{frame}}$ and $\mathcal{O}_{\mathrm{TOT}}^{\mathrm{TTI}}$) using the aforementioned detailed model lead to the peak data rates or throughputs represented in Figure 3.9 (a), where red columns represent the throughput computed during 1 TTI, $\mathcal{T}_{\mathrm{TOT}}^{\mathrm{TTI}}$, while blue columns represent the throughput computed during 1 frame, $\mathcal{T}_{\mathrm{TOT}}^{\mathrm{frame}}$. For comparative purposes, we have also included two results that usually appears in throughput estimations: 403.2 Mbps (the ideal limit when ECR = 1 and no overhead is used) and 326.4 Mbps (found in Table 13-1 in the Technical Specification 3GPP TR 25.912 [166], which do really introduce overhead but still maintains ECR = 1). Our limiting case $\mathcal{T}_{\mathrm{TOT}}^{\mathrm{TTI}} = 313.6$ Mbps is lower although more realistic because it assumes ECR < 1 and takes into account all overhead components. In this respect, focusing on our results, at first sight, in absolute terms (Mbps), $\mathcal{T}_{\mathrm{TOT}}^{\mathrm{TTI}}$ and $\mathcal{T}_{\mathrm{TOT}}^{\mathrm{frame}}$ reach similar results, and one could think that $\mathcal{T}_{\mathrm{TOT}}^{\mathrm{TTI}}$ and $\mathcal{T}_{\mathrm{TOT}}^{\mathrm{frame}}$ are interchangeable within all the LTE BW. However, to find out to what extent $\mathcal{T}_{\mathrm{TOT}}^{\mathrm{TTI}}$ and $\mathcal{T}_{\mathrm{TOT}}^{\mathrm{frame}}$ are "similar", it is convenient to compute some *relative metric*, which provides a more accurate insight.

This is why we use the *relative error*

$$\varepsilon_{\mathcal{T}_{\mathrm{TOT}}}(\%) = \frac{\mathcal{T}_{\mathrm{TOT}}^{\mathrm{TTI}} - \mathcal{T}_{\mathrm{TOT}}^{\mathrm{frame}}}{\mathcal{T}_{\mathrm{TOT}}^{\mathrm{TTI}}} \cdot 100, \tag{3.46}$$

which is the relative error made when computing the throughput at TTI level instead of at frame level. It has been represented in Figure 3.9 (b) as a function of BW (MHz). Note that at BW = 5 MHz, this error is as small as $\varepsilon_{\mathcal{T}_{\mathrm{TOT}}} = 1.57\%$. And even more interesting, in those scalable bandwidths for which LTE exhibits the highest data rates (BW $= 10, 15, 20$ MHz), the relative error is $\Delta\mathcal{T}_{\mathrm{TOT}} < 1\%$. The practical consequence is that if an *accuracy lower than* 1% is admitted, there is *no practical difference between computing the overhead either at TTI level or at frame level*: $\mathcal{T}_{\mathrm{TOT}}^{\mathrm{frame}} \approx \mathcal{T}_{\mathrm{TOT}}^{\mathrm{TTI}} = \mathcal{T}$.

As the rule of thumb, the system has its highest data-rate, when the highest allowable bandwidth is in-use (BW= 20 MHz), while applying the highest MCS order and data-rate enhancement techniques (64QAM with MIMO 4×4) to a single user, in order to minimize the overhead. However, the highest MCS and antenna multiplexing may still be interpreted with different values of ECR and Ω, and the

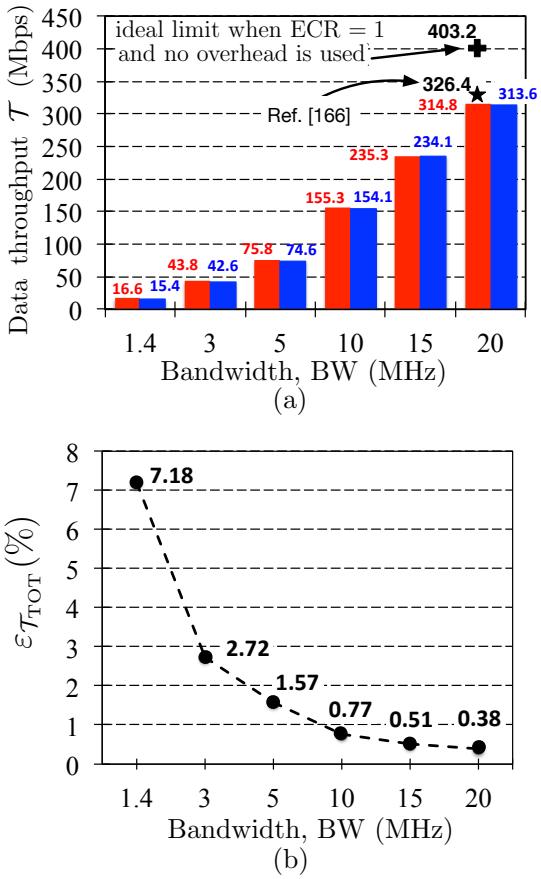


Figure 3.9: (a) Data throughput $\mathcal{T}$ (Mbps) as a function of BW (MHz) computed by using the parameter values listed in Table 3.10 and the overhead values shown in Figure 3.7. Red columns represent the throughput computed during 1 TTI, $\mathcal{T}_{\mathrm{TOT}}^{\mathrm{TTI}}$, while blue columns represent the throughput computed during 1 frame, $\mathcal{T}_{\mathrm{TOT}}^{\mathrm{frame}}$. (b) Relative error in throughput estimation, $\varepsilon_{\mathcal{T}_{\mathrm{TOT}}}(\%)$ (Expression 3.46), as a function of BW (MHz).

overhead considerations, such as overhead signals included in the calculation and data allocation period (Δt) will still result in different maximum data-rate for the system. The obtained maximum LTE downlink throughput values for above variables are compared in Table 3.11. These results and relatively high difference between the obtained data-rates once again highlight the need for an adequate estimation of the overhead and user demand modeling. This difference in throughput, varies from ideal case with no overhead, resulting in *403.2 Mbps*, to *326.4 Mbps* calculated value in 3GPP study [166] with ECR = 1 per TTI. The obtained data-rate will lessen if the ECR= 0.925 is considered according to highest CQI possible (CQI = 15) and the Δt for overhead calculation assumes full overheads in the $\Delta t = 1$ Frame instead of 1 TTI. The maximum data-rate drops to the least value of *240.8 Mbps* if instead

of multiplexing factor of $\Omega_{\max} = 4$, the spectral efficiency of $\Omega = 3.2$ is considered for 4×4 MIMO.

Table 3.11: Comparison of LTE maximum data-rate in DL with highest fixed values for: bandwidth (BW = 20 MHz); MCS order (64 QAM with $Q = 1$); deployed spacial multiplexing (MIMO 4×4), and variable values for overhead, ECR, Ω, and Δt.

ECR (Q=6, 64QAM)	Ω ($q = 1$, 4×4)	Δt (TTI or Frame)	Overhead	Data-rate (Mbps)
1	$\Omega_{\max}$=4	TTI	None	**403.2**
1	$\Omega_{\max}$=4	TTI	ICS = $1_{OFDMSymbol}$, $\mathcal{O}_{\text{RS}}$	**326.4**
0.925 (CQI=15)	$\Omega_{\max}$=4	TTI	$\mathcal{O}_{\text{TOT}}^{\text{TTI}}$	**314.6**
0.925 (CQI=15)	$\Omega_{\max}$=4	Frame	$\mathcal{O}_{\text{TOT}}^{\text{Frame}}$	**313.6**
0.925 (CQI=15)	$\Omega_{\max}$=4	TTI	Full ($ICS = 1_{OFDMSymbol}$, $\mathcal{O}_{\text{TOT}}^{\text{TTI}}$	**302.2**
0.925 (CQI=15)	$\Omega_{\max}$=4	Frame	Full ($ICS = 1_{OFDMSymbol}$, $\mathcal{O}_{\text{TOT}}^{\text{Frame}}$	**300.9**
0.925 (CQI=15)	Ω=3.2	Frame	Full (($ICS = 1_{OFDMSymbol}$, $\mathcal{O}_{\text{TOT}}^{\text{Frame}}$	**240.8**

To conclude this study case we would like to emphasize that this is the best one in the sense that minimizes the total overhead (optimum parameters in Table 3.6). However, there are other situations in which it is not possible to use all the optimum overhead parameters. This increase overhead and reduces the number of useful resources $\mathcal{U}_{\text{TOT}}^{\Delta t}$ will vary. Thus, the questions arising are to what extent they change and whether or not they have substantial influences on the throughput from an operative viewpoint. We begin this research by exploring the influence of N_g.

3.9.2 Exploring the influence of the Hybrid ARQ configuration

As mentioned, the PHICH carries the Hybrid-ARQ acknowledgment in DL (corresponding to its previous counterpart uplink data transfers). Its corresponding overhead $\mathcal{O}_{\text{PHICH}}^{\text{TTI}}$ depends, among others, on the PHICH Group Scaling Factor, N_g. Figure 3.10 will assist us to quantify the influence of N_g [162], whose possible values $\{1/6, 1/2, 1, 2\}$ determine the number of PHICH groups (Table 3.9).

Figure 3.10 (a) shows the throughput $\mathcal{T}_{\text{TOT}}^{\text{TTI}}$ (Mbps) as a function of BW (MHz) parametrized by $N_g = 1/6, 1/2, 1, 2$, respectively, while Figure 3.10 (b) represents the relative errors made with respect to the best case ($N_g = 1/6$):

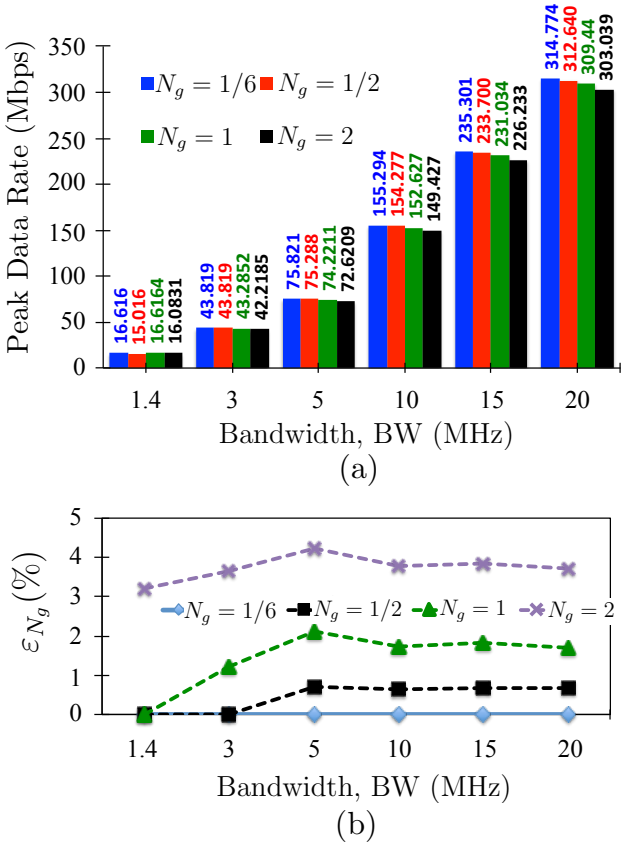


Figure 3.10: (a) Throughput (Mbps) as a function of BW (MHz) parametrized by $N_g = 1/6, 1/2, 1, 2$, respectively. (b) Throughput relative error $\varepsilon_{N_g}(\%)$ defined in Expression 3.47.

$$\varepsilon_{N_g}(\%) \doteq \frac{\mathcal{T}_{\text{TOT}}|_{(N_g=1/6)} - \mathcal{T}_{\text{TOT}}|_{N_g}}{\mathcal{T}_{\text{TOT}}|_{(N_g=1/6)}} \cdot 100 \tag{3.47}$$

Note that $\varepsilon_{N_g=2} \lesssim 4\%$. That is, working with the case in which $N_g = 2$ (a worst case that maximizes its corresponding overhead) only reduces the peak bit rate in $\approx 4\%$ when compared to that computed when using $N_g = 1/6$ (best case). This worst case leads to a bit rate 303 Mbps (using $N_g = 1/6$). If we consider as acceptable a maximum error of 4%, this will provide us a a safety factor that includes the possibility of using any allowed value of N_g. This is why we consider henceforth $N_g = 2$ since it is a conservative design (worst case) that, from a practical point of view, underestimate very slightly the peak-data rate.

3.9.3 Influence of MAT and resource allocation types

As described in Section 3.3.2, multi-antenna technologies allows for increasing throughput. These techniques depend on the number of transmitting antennae (n),

number of layers (M), and the number of transmitted codewords q. As summarized in Table 3.4, they determine the value of Ω, related to the number of streams. In turn, the scheduling type and DCI format to be used are related to the multi-antennae and multi-layer transmission scheme selected (Table 3.5).

Regarding these interrelated factors, Figure 3.11 shows the peak data rate (Mbps) as a function of BW (MHz), parametrized by n and the allowed values for q in each configuration. This leads to the five possible cases represented: 1) SISO, with n =1 and $q = 0$. 2) 2×2 MIMO and $q = 0$. 3) 2×2 MIMO and $q = 1$. 4) 4×4 MIMO and $q = 0$. 5) 4×4 MIMO and $q = 1$. As has been justified in Subsection 3.9.2, we have assumed the worst case $N_g = 2$ since it leads to a bound with $\varepsilon_{N_g=2} \lesssim 4\%$. Figure 3.11 shows that:

1. For $q = 1$ (that is, using two codewords), introducing MIMO schemes with $q = 1$ –green and blue bars– leads to superior performance than that of SISO ($n = 1$) for any BW value.

2. However, for $q = 0$ (using only 1 codeword), introducing MIMO schemes –red and yellow bars– leads to *worse* throughputs. For instance, if we focus on those results marked by ★ symbol for BW = 20 MHz, the peak data rate for SISO ($n = 1$, grey bar) is $\approx$ 85 Mbps, which is greater than the one of 2×2 MIMO with $q = 0$ ($\approx$ 82 Mbps, red bar), and 4×4 MIMO with $q = 0$ ($\approx$ 77 Mbps, yellow bar). This is because of the additional reference signal resources that are imposed by higher order transmission schemes, which increase the percentage of overhead RE with respect to the available ones. This has sense because, as stated by Expressions (3.17) and (3.18), once the BW has been selected (and thus N_{RB}), $N_{\mathrm{RS}|(1\ \mathrm{TTI})} = 8$, 16 and 24 for $n = 1, 2, 4$, respectively.

For completeness, Figure 3.12 shows the the way the different overhead *components* have a *different impact* on the total overhead as a function of n, and reveals:

1. In the three transmitting configurations, for BW $\geq$ 5 MHz, the dominant overhead components are, at great extent, $\mathcal{O}_{\mathrm{RS}}$ and, to a lesser extent, $\mathcal{O}_{\mathrm{PHICH}}$.

2. For BW $<$ 5, the overhead components exhibit different behaviors, which are not easy to generalize. The effects are particularly clear at BW = 1.4 MHz. As n reduces –when going from Figure 3.12 (a) to figure (c)–, the influence of $\Delta\mathcal{O}_{\mathrm{PDCCH}}$ increases in the detriment of $\Delta\mathcal{O}_{\mathrm{RS}}$:

   - For $n = 4$, $\Delta\mathcal{O}_{\mathrm{RS}} \approx 52\%$ and $\Delta\mathcal{O}_{\mathrm{PDCCH}} \approx 25\%$ (out of the total overhead)
   - For $n = 2$, $\Delta\mathcal{O}_{\mathrm{RS}}$ reduces up to 42% while $\Delta\mathcal{O}_{\mathrm{PDCCH}}$ rises up to $\approx 30\%$
   - For $n = 1$, $\Delta\mathcal{O}_{\mathrm{PDCCH}}$ becomes clearly dominant ($\approx 38\%$) when compared to the others.

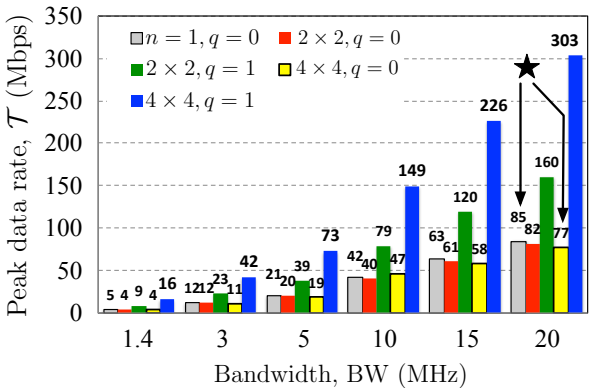


Figure 3.11: Lower bound ($N_g = 2$) of the peak data rate (Mbps) as a function of BW (MHz), parametrized by n and q: 1) SISO, with $n = 1$ and $q = 0$. 2) 2×2 MIMO with $q = 0$. 3) 2×2 MIMO with $q = 1$. 4) 4×4 MIMO with $q = 0$. 5) 4×4 MIMO with $q = 1$.

3.9.4 Maximum number of UE per TTI, $N_{\text{UE}}^{\text{TTI}_i}|_{\text{MAX}}$

As explained in Section 3.7, the computing overhead methodology presented in this paper is also useful to compute the maximum number of UEs that can be scheduled in each TTI. This feature is different from the maximum number of users' equipments supported by a cell, as the active users can be scheduled in distributed TTIs.

Based on Expression (3.36), Figure 3.13 shows the maximum number of users per TTI, $N_{\text{UE}}^{\text{TTI}_i}|_{\text{MAX}}$, parametrized by $N_g = 1/6, 1/2, 1, 2$, as a function of BW, for different configurations: SISO (a); 2×2 MIMO with $q = 0$ (b) and $q = 1$ (c); 4×4 MIMO with $q = 0$ (d) and $q = 1$ (e). Note that in SISO we have considered only $q = 0$ because in SISO only 1 codeword is allowed.

Aiming at explaining the influence of parameter variation in Figure 3.13, the following discussion is based on increasing the value of one parameter (for instance, n) while those of the other two are kept constant (BW, and q). The process is repeated to analyze all the parameters involved. We use $\uparrow$ symbol to represent such an increment. With this in mind, Figure 3.13 shows that:

- Increasing the number of antennae n reduces the maximum number of users per TTI, $N_{\text{UE}}^{\text{TTI}_i}|_{\text{MAX}}$, because it requires a higher $\mathcal{O}_{\text{RS}}^{\text{TTI}}$. This is evident when comparing (a), (b) and (d). For instance, for $N_g = 2$, $N_{\text{UE}}^{\text{TTI}_i}|_{\text{MAX}}$ reduces from 21 users per TTI (SISO, in Subfigure (a)) to 17 users per TTI (4×4 MIMO and $q = 0$, in Subfigure (d)). This has sense because as $n \uparrow \Rightarrow \uparrow \mathcal{O}_{\text{RS}}^{\text{TTI}}$ (Expression (3.18)) $\Rightarrow \uparrow \Gamma$ (see Expression (3.30)) $\Rightarrow N_{\text{UE}}^{\text{TTI}_i}|_{\text{MAX}} \downarrow$ (as stated by Expression (3.18)).

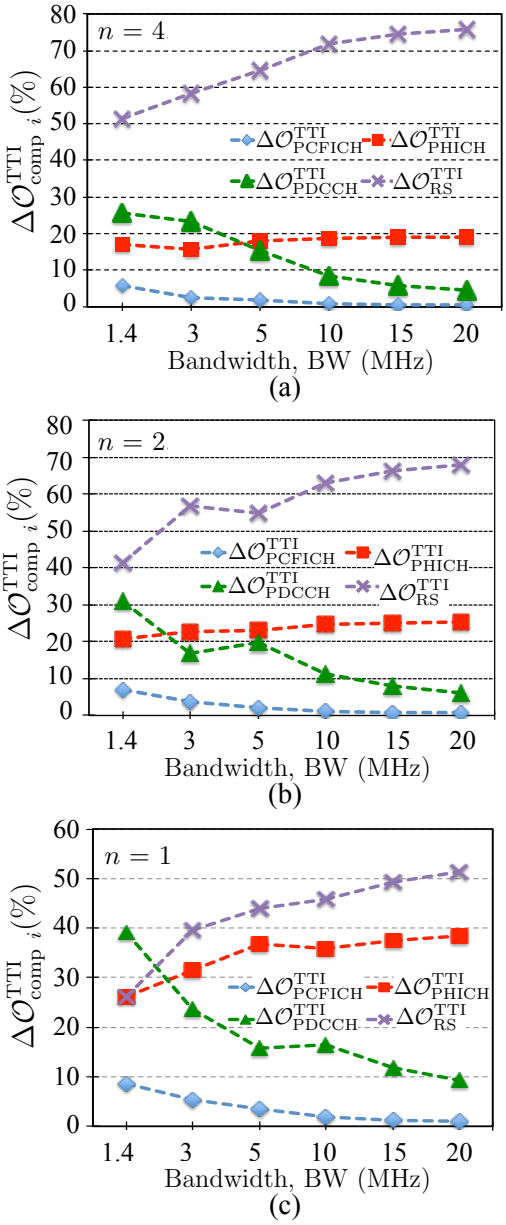


Figure 3.12: Overhead percentage (%), with respect to the total REs, as a function of BW, for $n = 4$ (a), $n = 2$ (b), and $n = 1$ (c).

- Increasing N_g reduces $N_{\mathrm{UE}}^{\mathrm{TTI}_i}|_{\mathrm{MAX}}$. This is clear, for instance, in Figure 3.13 (c), for BW= 20 MHz. This is because if $N_g \uparrow \Rightarrow \uparrow \mathcal{O}^{\mathrm{TTI}}_{\mathrm{PHICH}}$ (see Expression 3.21 and Table 3.9) $\Rightarrow \uparrow \Gamma \Rightarrow$ (Expression (3.30)) $\Rightarrow N_{\mathrm{UE}}^{\mathrm{TTI}_i}|_{\mathrm{MAX}} \downarrow$

- Increasing the number of codewords via q reduces $N_{\mathrm{UE}}^{\mathrm{TTI}_i}|_{\mathrm{MAX}}$. This is because if $q \uparrow$ then $\uparrow \mathcal{O}^{\mathrm{TTI}}_{\mathrm{PHICH}} \Rightarrow \uparrow \Gamma \Rightarrow N_{\mathrm{UE}}^{\mathrm{TTI}_i}|_{\mathrm{MAX}} \downarrow$.

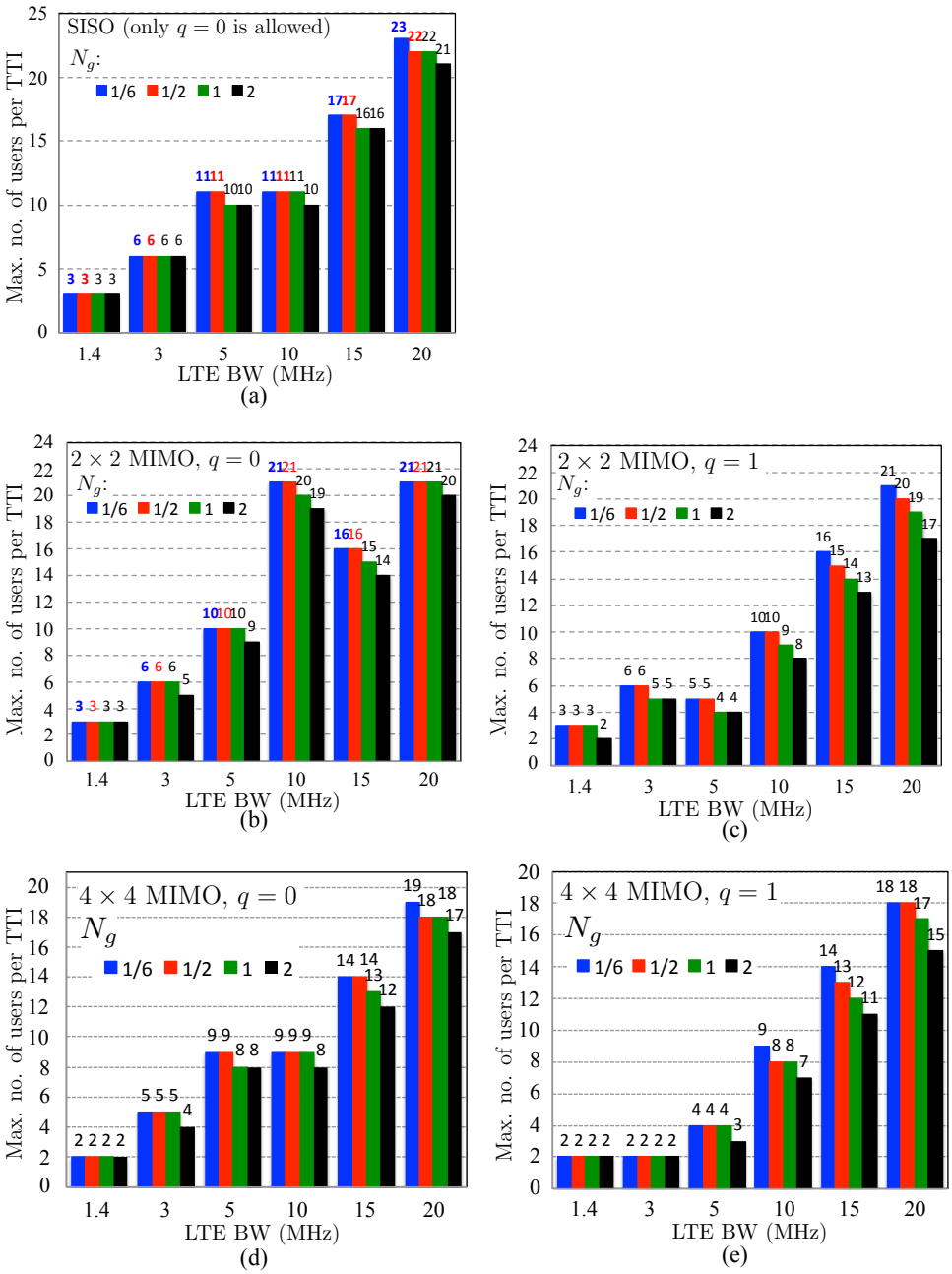


Figure 3.13: Maximum number of users per TTI, parametrized by $N_g = 1/6, 1/2, 1, 2$, as a function of BW for different configurations: SISO (a); 2×2 MIMO with $q = 0$ (b) and $q = 1$ (c); 4×4 MIMO with $q = 0$ (d) and $q = 1$ (e).

3.9.5 Multiple users with different services

Table 3.12 summarizes the transmission and reception conditions of the simulated scenario. It corresponds to a system with BW = 20 MHz in which there are 10 users with different Multiple Antenna Technologies (MAT). These 10 users have to share $N_{\mathrm{RB}} = 100$ RBs corresponding to BW = 20 MHz (see Table 3.3). It has been assumed that users are randomly distributed so that any UE_j (with $j = 1, 2, \cdots, 10$) has a given CQI value, which, using Table 3.2, allows for obtaining the corresponding values for Q and ECR.

Table 3.12: Values of the parameters involved in the multi-user scenario simulated. ID means identifier. MAT stands for Multiple Antenna Technology.

User ID	MAT	Ω	CQI	Q	ECR
1	4×4 MIMO	4	6	2	0.588
2	2×2 MIMO	2	9	4	0.601
3	SISO	1	15	6	0.926
4	4×4 MIMO	4	10	6	0.455
5	2×2 MIMO	2	9	4	0.601
6	4×4 MIMO	4	5	2	0.438
7	SISO	1	12	6	0.650
8	2×2 MIMO	2	8	4	0.478
9	4×4 MIMO	4	11	6	0.554
10	4×4 MIMO	4	14	6	0.825

Based on the user's data listed in Table 3.12, we have applied the proposed method to compute the corresponding throughput in two different cases. These differ in the way the available resource blocks are assigned to the ten users. This can be done by using the parameter w_j defined by Expression 3.6, which is just the fraction of resource blocks assigned to each user. In the first case study we have considered that all users received the same fraction of RBs, that is $w_1 = w_2 = \cdots = w_{10} = 0.1$. The corresponding throughout for each user has been represented by means of the blue bars in Figure 3.14 (a). The second case study corresponds to a resource allocation strategy that aims to ensure that all users have approximately the same bit rate (red bars in Figure 3.14 (a)). This is done by adjusting each $w_j = N_{\mathrm{RB}}^{\mathrm{UE}_j}/N_{\mathrm{RB}}$. The corresponding number of RBs assigned to each user $N_{\mathrm{RB}}^{\mathrm{UE}_j}$ has been represented in Figure 3.14 (b). Note that the first strategy, which distributes resources equally among all users, leads to a service scenario in which there are users with high bit rate (for instance, user 10) in the detriment of others with much lower bit rate (see users 1 and 2). This is because of the different reception conditions (different CQI values) and different MAT schemes (SISO vs. 4×4 MIMO, for instance). The second strategy aims to compensate for these inequalities and help all users have a DL bit rate with mean value = 6.927336 Mbps and variance = 998 bps.

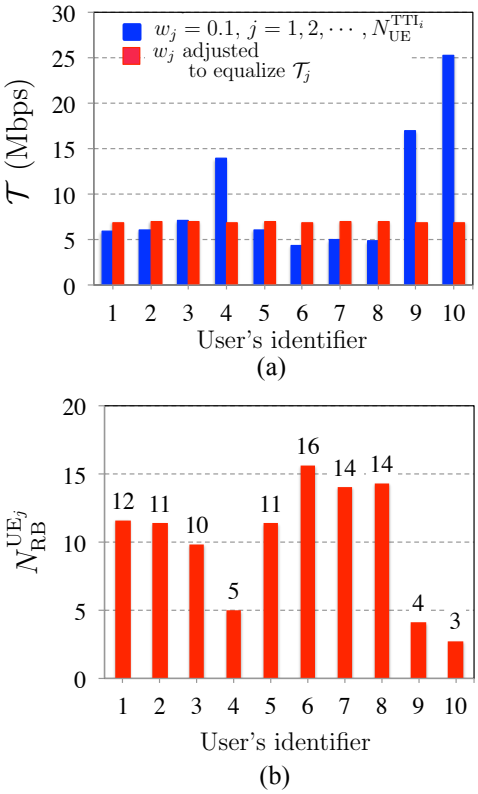


Figure 3.14: (a) Throughput $\mathcal{T}$ (Mbps) of each of the ten users whose DL conditions have been listed in Table 3.12. Blue columns represent each user's $\mathcal{T}_j$ when $w_j = 0.1$, while red columns show $\mathcal{T}_j$ when coefficients w_j are adjusted aiming at equalize each users' $\mathcal{T}_j$ (mean value = 6.927336 Mbps, variance = 998 bps). (b) Corresponding number of RBs that each user UE_j receives during TTI_i, $N_{\mathrm{RB}}^{\mathrm{UE}_j}$ (see Expression (3.6)) when coefficients w_j are adjusted aiming at equalizing $\mathcal{T}_j$.

3.10 Summary and conclusions

Overhead ($\mathcal{O}$) Resource Elements (REs) in the downlink (DL) of Long Term Evolution (LTE) networks, which are necessary for controlling, signaling and synchronization tasks at the Physical (PHY) level and Media Access Control (MAC) sub-level, play a key role to accurately compute LTE performance (basically, data throughput $\mathcal{T}$) in a variety of possible configurations, despite not having received much interest in the technical literature to the point that, to the best of our knowledge, there is *no* complete model containing all components and dependencies.

Just in this respect, this work has presented a *data-rate* based *performance model* that aims at accurately and *dynamically* computing all the overhead REs (and, thus the real throughput $\mathcal{T}$) within a unified framework, which takes into account all the possible variables it depends on: channel bandwidth (BW), resource allocation type,

multiple antenna technology (MAT), Hybrid Automatic Repeat Request (HARQ) configuration, and number and reception conditions of user equipments (UEs). The model includes all the mechanisms that generate overhead: Physical Downlink Control Channel Overhead ($\mathcal{O}^{\Delta t}_{\text{PDCCH}}$), Reference Signal Overhead ($\mathcal{O}^{\Delta t}_{\text{RS}}$), Physical Hybrid ARQ Indicator Channel Overhead ($\mathcal{O}^{\Delta t}_{\text{PHICH}}$), Physical Control Format Indicator Channel Overhead ($\mathcal{O}^{\Delta t}_{\text{PCFICH}}$), Synchronization Signal Overhead ($\mathcal{O}^{\Delta t}_{\text{PSCH}}$) and Broadcast Channel Overhead ($\mathcal{O}^{\Delta t}_{\text{PBCH}}$). The last two are repeated only once per frame (= 10 ms), while the first four are repeated once per subframe or Transmission Time Interval (TTI), the smallest time interval (= 1 ms) in which LTE can assign resources to users in a cell. Δt stands for the time period over which overhead is calculated depending on user's service. Computing overhead at TTI scale ($\Delta t = 1$ TTI) is specially practical in applications using Transmission Control Protocol (TCP), in which user's traffic is bursty and dynamic. However, there are services, such as VoIP, that require persistent scheduling for allocating resources to a user for a time period longer than one TTI ($\Delta t = p$ TTI, p being an integer). The model allows thus for computing the total overhead at both TTI and frame scales, $\mathcal{O}^{\text{TTI}}_{\text{TOT}}$ and $\mathcal{O}^{\text{frame}}_{\text{TOT}}$, respectively, although it is also possible to compute $\mathcal{O}^{p\ \text{TTI}}_{\text{TOT}}$ under a persistent scheduling for a time period of p TTIs.

Such versatility (in computing overhead over the required number of TTIs) and flexibility (including all possible variables) made the model be easily adaptable for all the different LTE system configurations and realistic situations. In this respect, the model has assisted us in: 1) accurately computing the contribution of *each* overhead mechanism and its corresponding influence on real data throughput; 2) optimizing throughput (by selecting the configurations that minimizes overhead) in each situation; 3) calculating the maximum number of users per TTI (which are restricted by a limited number of controlling REs); and 4) dynamically computing throughputs, not only in single user scenarios, but also in multi-service and multi-user scenarios with adaptive MAT for each user. In our experimental work we have explored the influence of the variables that overhead depends on, the main results being as follow.

1. Influence of BW

The first result in this respect is that the *relative* influence of overhead (the percentage with respect to the total available resources) *reduces* as BW increases. Specifically, in the optimum 4×4 MIMO configuration, for a channel bandwidth BW = 20 MHz, the percentage (with respect to the total *available resources*) of both computed overheads, $\Delta\mathcal{O}^{\text{frame}}_{\text{TOT}}$ and $\Delta\mathcal{O}^{\text{TTI}}_{\text{TOT}}$, tend to an upper bound of about 15.9%. While, for BW = 1.4 MHz, $\Delta\mathcal{O}^{\text{frame}}_{\text{TOT}}\ |_{(\text{BW=1.4 MHz})} = 31.1\%$ is appreciably higher than $\Delta\mathcal{O}^{\text{TTI}}_{\text{TOT}}\ |_{(\text{BW=1.4MHz})} = 25.4\%$, however, as BW increases, their respective values reduce, both being below 17% for BW $= 10, 15, 20$ MHz. The results from [66, 72] underestimate overhead because they do not take into account all the overhead components: 12.12% [66] compared to our (total) result $\Delta\mathcal{O}^{\text{TTI}}_{\text{TOT}}\ |_{(\text{BW=10MHz})} = 15.8\%$, and 14.3% [72] compared to our result $\Delta\mathcal{O}^{\text{TTI}}_{\text{TOT}}\ |_{(\text{BW=20MHz})} = 15.6\%$.

A second interesting block of results is related to the percentage of each com-

ponent over the total overhead $\mathcal{O}_{\mathrm{TOT}}^{\mathrm{TTI}}$. The results show that $\mathcal{O}_{\mathrm{RS}}^{\mathrm{TTI}}$ plays the key role since is much higher than all the other components in all the LTE bandwidths. Furthermore, as BW increases, $\mathcal{O}_{\mathrm{RS}}^{\mathrm{TTI}}$ also increases, while the percentage of others reduce. For BW = 20MHz, this element is as large as $\Delta\mathcal{O}_{\mathrm{RS}}^{\mathrm{TTI}} = 91\%$ of all counted overheads. Moreover, $\Delta\mathcal{O}_{\mathrm{RS}}^{\mathrm{TTI}} + \Delta\mathcal{O}_{\mathrm{PDCCH}}^{\mathrm{TTI}}$ ranges between 84% and 97% in all the BW range 1.4 MHz $\leq$ BW $\leq$ 20 MHz. When applying the model at frame scale, the results exhibit the same trends, although the relative contributions of $\mathcal{O}_{\mathrm{RS}}^{\mathrm{frame}}$ and $\mathcal{O}_{\mathrm{PDCCH}}^{\mathrm{frame}}$ are slightly smaller. This is because, at frame scale, there are overhead components ($\mathcal{O}_{\mathrm{SCH}}^{\mathrm{frame}}$ and $\mathcal{O}_{\mathrm{PBCH}}^{\mathrm{frame}}$) that do *not* appear in a generic TTI, reducing thus the relative contribution of $\mathcal{O}_{\mathrm{RS}}^{\mathrm{frame}}$ and $\mathcal{O}_{\mathrm{PDCCH}}^{\mathrm{frame}}$, which, in any case, are clearly the dominant contributions to the total overhead.

The relative error made when computing the throughput at TTI level ($\mathcal{T}_{\mathrm{TOT}}^{\mathrm{TTI}}$) instead of at frame level ($\mathcal{T}_{\mathrm{TOT}}^{\mathrm{frame}}$), $\Delta\mathcal{T}_{\mathrm{TOT}}$, has been found to be $\Delta\mathcal{T}_{\mathrm{TOT}} < 1\%$ in those channel bandwidths for which LTE exhibits the highest data rates (BW $= 10, 15, 20$ MHz). The practical consequence is that if an accuracy lower than 1% is admitted, there is *no* practical difference between computing the overhead either at TTI level or at frame level: $\mathcal{T}_{\mathrm{TOT}}^{\mathrm{frame}} \approx \mathcal{T}_{\mathrm{TOT}}^{\mathrm{TTI}} = \mathcal{T}$.

Finally, as all overhead mechanisms have been computed in detail, the real data throughput results is being slightly lower than those computed using less realistic models. In particular, the limiting value has been found to be $\mathcal{T} = 313.6$ Mbps.

2. Influence of the Hybrid ARQ configuration

The Physical Hybrid ARQ Indicator Channel (PHICH) carries the Hybrid-ARQ acknowledgment in DL (corresponding to its previous counterpart uplink data transfers). We have shown that its corresponding overhead $\mathcal{O}_{\mathrm{PHICH}}^{\mathrm{TTI}}$ depends, among others, on the PHICH Group Scaling Factor, N_g, which can only take values in the set $\{1/6, 1/2, 1, 2\}$. We have found that working with the case in which $N_g = 2$ (a worst case that maximizes $\mathcal{O}_{\mathrm{PHICH}}^{\mathrm{TTI}}$) only reduces the peak bit rate in $\approx 4\%$ when compared to that computed when using $N_g = 1/6$ (best case). This worst case leads to a bit rate 303 Mbps. If we consider as acceptable a maximum error of 4%, this will provide us a a safety factor that includes the possibility of using any allowed value of N_g.

3. Influence of MAT and resource allocation type

Multiple antennas technologies impact is classified into: Diversity that improves the UE's reception condition by duplicating the streams, thus allowing the use of higher order MCS; as well as MIMO, that allows sending parallel streams which allows for increasing throughput. The increase depends on the number of transmitting (and reception) antennae (n), the number of layers (M), and the number of transmitted codewords (q). The resource allocation types defined in LTE lead to a number of allowed feasible combinations of values for n, M and q. Specifically, we have computed the peak data $\mathcal{T}$ (Mbps) as a function of BW (MHz), parametrized by n and the allowed values for q in five possible cases: 1) SISO (Single-Input and Single-Output, $n = 1$), with n =1 and $q = 0$ (only 1 codeword is allowed); 2) 2×2 MIMO (Multiple-Input and Multiple-Output) and $q = 0$; 3) 2×2 MIMO and $q = 1$;

4) 4×4 MIMO and $q = 0$; and 5) 4×4 MIMO and $q = 1$.

The first interesting result is that increasing the number of antennas n in MIMO schemes *not always* leads to a higher throughput. When using 2 codewords ($q = 1$), increasing n does lead to a superior performance than that of SISO for any BW value. However, when using only 1 codeword ($q = 0$), increasing n does *not* increase throughput but decreases it. For instance, for BW = 20 MHz, the peak data rate for SISO is $\mathcal{T}_{\text{SISO}} \approx 85$ Mbps, which is greater than the one of 2×2 MIMO with $q = 0$ ($\mathcal{T}_{2\times2\ \text{MIMO}} \approx 82$ Mbps), and 4×4 MIMO with $q = 0$ ($\mathcal{T}_{4\times4\ \text{MIMO}} \approx 77$ Mbps). This is because of the additional reference signal resources required for higher order transmission schemes, which increase the percentage of overhead RE with respect to the available ones.

The second interesting result is related to the relative weight of each overhead component (with respect to the total) as a function of $n = 1, 2, 4$. In the three transmitting configurations, for BW ≥ 5 MHz, the dominant overhead components are, at great extent, $\mathcal{O}_{\text{RS}}$ and, to a lesser extent, $\mathcal{O}_{\text{PHICH}}$. However, at BW = 1.4 MHz, as n reduces, the influence of $\mathcal{O}_{\text{PDCCH}}$ increases in the detriment of $\mathcal{O}_{\text{RS}}$: for $n = 4$, $\Delta\mathcal{O}_{\text{RS}} \approx 52\%$ and $\Delta\mathcal{O}_{\text{PDCCH}} \approx 25\%$, while, for $n = 1$, $\mathcal{O}_{\text{PDCCH}}$ becomes clearly dominant ($\Delta\mathcal{O}_{\text{PDCCH}} \approx 38\%$) when compared to the others.

4. Maximum number of users per TTI

The model includes the existence of a maximum number of users per TTI based on the fact that there is a limited number of controlling OFDM symbols for shared-data transmitting subframes. Specifically, we have computed the maximum number of users per TTI, $N_{\text{UE}}^{\text{TTI}_i}|_{\text{MAX}}$, parametrized by $N_g = 1/6, 1/2, 1, 2$, as a function of BW, for different configurations: SISO with $q = 0$; 2×2 MIMO with $q = 0$ and $q = 1$; and 4×4 MIMO with $q = 0$ and $q = 1$. The main results are: 1) Increasing the number of antennae n reduces the maximum number of users per TTI, $N_{\text{UE}}^{\text{TTI}_i}|_{\text{MAX}}$, because it requires higher $\mathcal{O}_{\text{RS}}^{\text{TTI}}$. 2) Increasing either N_g or q reduces $N_{\text{UE}}^{\text{TTI}_i}|_{\text{MAX}}$ because $\mathcal{O}_{\text{PHICH}}^{\text{TTI}}$ increases.

Finally, the versatility (in computing overhead over the required number of TTIs for each user's service) and the flexibility (including all possible variables) of our method allows for *optimizing throughput* (the metric for performance evaluation), by selecting the configuration that minimizes overhead in any LTE operation condition. Then these results can be tabulated and easily programmed to be used in *real-time*. Thus, the model is useful for dynamic resource allocation in multi-user scenarios with multi-antenna transmission schemes since it is able to include different resource allocation strategies by simply varying the amount of resource blocks assigned to each user. In the end it could be applied in more detailed system performance evaluation studies, such as capacity estimation, cell and network planning along with service provision design and optimization.

Chapter 4

Conclusions and Future Work

4.1 Summary and conclusions

When we began this thesis in 2012 there was two main technologies competing for the International Mobile Telecommunications (IMT)-Advanced initiative, WiMAX and LTE, and it was not clear at that moment which of the two would win and lead the Broadband Mobile Internet [56].

However, at present, LTE is the undisputed winner for broadband mobile cellular networks, while WiMAX has now been confined to very specific market niches such as aviation industry [167–170], smart grids [171–173], smart city [174], or point-to-point long distance and high rate radio transmission in applications such as oil and gas companies. See [175] for further details.

In any case, what we do have in common in both standards is that LTE and Mobile-WiMAX are both very complex solutions that, apart from technological similarities, such as use of OFDM technology, have many structural and implementation differences. Additionally, the standards that define these two broadband mobile solutions, both contain numerous flexibilities. The configuration choices are left open for the vendors to find their optimal solution according to their service area. The performance of these networks is therefore highly dependent on solving this complex optimized design problem i.e. system vs. load. The network planning and optimize design in new generation mobile networks is an active research line, specially in LTE, the clear winner. High-level and at the same time accurate performance modeling are of high value in planning cellular networks, as full scale dynamic simulations are not affordable in time and complication for large planning scenarios.

The objective of this thesis has been to provide a Performance Key Indicator (KPI) for 4G broadband mobile technologies, defined as "The maximum number of simultaneous multi-profile users that are jointly supported with each specific access point, under specific system configuration". This KPI has been found to be a representation of the network operational capacity and corresponds to the goodput of the system at full-load. The obtained capacity can be used as the basis for network design, planning and dimensioning purposes, as well as cost estimation of the new

generation cellular networks.

We have explored this research goal in a unified framework using:

- System Configuration: The technological features of the 4G solution (i.e. LTE or Mobile-WiMAX) have been considered in detail with a change sensitive focus for different system configuration options. The available resources for each configuration and allocation of these resources have been modeled:
  - As the number of users increases, the operational system overheads to keep track of these users also increase. So apart from the fixed system overheads and data load consumed by the users, a dynamic overhead calculation is required to know the remaining resources at each moment.
  - The resource allocation mechanisms of the technology under study needs to be mathematically formulated to obtain the available resources below each specific configuration and to know how each technology shares these available resources between its end users.
- Service Profile: The end users that populate the service area, each may be placed at different location and may request different sets of data services. Different distribution of the users in service area results in different signal reception conditions, and different traffic services that they request may consume different amount of system resources. Thus it is essential to elaborate a Service Profile definition to model the user's condition and demand.
- System Capacity: An interactive algorithm is required to compare the system available resources with the system demand at each instance, to determine if the resources can meet new demands, and if not, present the maximum affordable demand as the system operational capacity.

With these considerations in mind, in Chapter 2 we have proposed a dynamic capacity estimation methodology for Mobile WiMAX systems, while the sector under study is simultaneously supporting multi-service users. Our algorithm contains two major blocks that are solved at each given time and in parallel in order to, on one hand, calculate the optimal available resource with respect to MCS distribution and multi-user burst construction and, from the other hand, calculate the minimum required resource consumption with respect to application profile and the multi-service data-rate requirements. The general formulations for overhead estimation are provided for any desired system parameters. The data-rate requirements of the covered multi-services are grouped into guaranteed and non-guaranteed data-rates and an application profile is defined as a consistent input for capacitating and dimensioning studies evaluation. Simulations are run with example values for MCS distribution and daily averaged application profile. The simulation results with regard to PHY+MAC overhead removal methodology underline the significant roll of overhead calculations in capacity estimation studies. The elaborated statistical application profile replaces the complexity of simulating a detailed scheduling strategy,

thus providing reference measures in order to evaluate the performance of different scheduling techniques in real time. Having the number of supported user with each WiMAX hotspot, with regards to the parameters that characterize the access point and the service area, the network planners can optimize the service provision in the coverage area. Furthermore, the proposed algorithm is highly change sensitive. In other words, the algorithm can be simply configured for each set of input parameters, in order to study the impact of each parameter change on system performance. The performance is evaluated in a single output result, that is, the number of supported multi-users (or the cell goodput). Further studies shall be done in order to improve the proposed algorithm with more implementation details, while keeping the simple structure of the proposed approach for capacity estimation. In continuation to this work, we plan to propose solutions in order to integrate the performance enhancing techniques in proposed algorithm and to include the short-scale channel state variations factor in resource allocation. Furthermore, the impacts of handover and mobility overheads on the Mobile WiMAX capacity estimation are other topics that will be discussed in our related future works.

Following the same scientific approach, Chapter 3 has presented a *data-rate* based *performance model* that aims at accurately and *dynamically* computing all the overhead Resource Elements (REs). Overhead ($\mathcal{O}$) REs in the downlink (DL) of Long Term Evolution (LTE) networks, which are necessary for controlling, signaling and synchronization tasks at the Physical (PHY) level and Media Access Control (MAC) sub-level, play a key role to accurately compute LTE performance (basically, data throughput $\mathcal{T}$) in a variety of possible configurations, despite not having received much interest in the technical literature to the point that, to the best of our knowledge, there is *no* complete model containing all components and dependencies.

Just in this respect, this Chapter has presented a *data-rate* based *performance model* that aims at accurately and *dynamically* computing all the overhead REs (and, thus the real throughput $\mathcal{T}$) within a unified framework, which takes into account all the possible variables it depends on: channel bandwidth (BW), resource allocation type, multiple antenna technology (MAT), Hybrid Automatic Repeat Request (HARQ) configuration, and number and reception conditions of user equipments (UEs). The model includes all the mechanisms that generate overhead: Physical Downlink Control Channel Overhead ($\mathcal{O}^{\Delta t}_{\mathrm{PDCCH}}$), Reference Signal Overhead ($\mathcal{O}^{\Delta t}_{\mathrm{RS}}$), Physical Hybrid ARQ Indicator Channel Overhead ($\mathcal{O}^{\Delta t}_{\mathrm{PHICH}}$), Physical Control Format Indicator Channel Overhead ($\mathcal{O}^{\Delta t}_{\mathrm{PCFICH}}$), Synchronization Signal Overhead ($\mathcal{O}^{\Delta t}_{\mathrm{PSCH}}$) and Broadcast Channel Overhead ($\mathcal{O}^{\Delta t}_{\mathrm{PBCH}}$). The last two are repeated only once per frame (= 10 ms), while the first four are repeated once per subframe or Transmission Time Interval (TTI), the smallest time interval (= 1 ms) in which LTE can assign resources to users in a cell. Δt stands for the time period over which overhead is calculated depending on user's service. Computing overhead at TTI scale ($\Delta t = 1$ TTI) is specially practical in applications using Transmission Control Protocol (TCP), in which user's traffic is bursty and dynamic. However, there are services, such as VoIP, that require persistent scheduling for allocating resources to a user for a time period longer than one TTI ($\Delta t = p$ TTI, p being an

integer). The model allows thus for computing the total overhead at both TTI and frame scales, $\mathcal{O}_{\mathrm{TOT}}^{\mathrm{TTI}}$ and $\mathcal{O}_{\mathrm{TOT}}^{\mathrm{frame}}$, respectively, although it is also possible to compute $\mathcal{O}_{\mathrm{TOT}}^{p\ \mathrm{TTI}}$ under a persistent scheduling for a time period of p TTIs.

Such versatility (in computing overhead over the required number of TTIs) and flexibility (including all possible variables) made the model be easily adaptable for all the different LTE system configurations and realistic situations. In this respect, the model has assisted us in: 1) accurately computing the contribution of *each* overhead mechanism and its corresponding influence on real data throughput; 2) optimizing throughput (by selecting the configurations that minimizes overhead) in each situation; 3) calculating the maximum number of users per TTI (which are restricted by a limited number of controlling REs); and 4) dynamically computing throughputs, not only in single user scenarios, but also in multi-service and multi-user scenarios with adaptive MAT for each user.

In our experimental work we have explored the influence of the variables that overhead depends on, the main results being as follow.

1. Influence of BW

The first result in this respect is that the *relative* influence of overhead (the percentage with respect to the total available resources) *reduces* as BW increases. Specifically, in the optimum 4×4 MIMO configuration, for a channel bandwidth BW = 20 MHz, the percentage (with respect to the total *available resources*) of both computed overheads, $\Delta\mathcal{O}_{\mathrm{TOT}}^{\mathrm{frame}}$ and $\Delta\mathcal{O}_{\mathrm{TOT}}^{\mathrm{TTI}}$, tend to an upper bound of about 15.9%. While, for BW = 1.4 MHz, $\Delta\mathcal{O}_{\mathrm{TOT}}^{\mathrm{frame}}\ |_{(\mathrm{BW=1.4\ MHz})} = 31.1\%$ is appreciably higher than $\Delta\mathcal{O}_{\mathrm{TOT}}^{\mathrm{TTI}}\ |_{(\mathrm{BW=1.4MHz})} = 25.4\%$, however, as BW increases, their respective values reduce, both being below 17% for BW $= 10, 15, 20$ MHz. The results from [66, 72] underestimate overhead because they do not take into account all the overhead components: 12.12% [66] compared to our (total) result $\Delta\mathcal{O}_{\mathrm{TOT}}^{\mathrm{TTI}}\ |_{(\mathrm{BW=10MHz})} = 15.8\%$, and 14.3% [72] compared to our result $\Delta\mathcal{O}_{\mathrm{TOT}}^{\mathrm{TTI}}\ |_{(\mathrm{BW=20MHz})} = 15.6\%$.

A second interesting block of results is related to the percentage of each component over the total overhead $\mathcal{O}_{\mathrm{TOT}}^{\mathrm{TTI}}$. The results show that $\mathcal{O}_{\mathrm{RS}}^{\mathrm{TTI}}$ plays the key role since is much higher than all the other components in all the LTE bandwidths. Furthermore, as BW increases, $\mathcal{O}_{\mathrm{RS}}^{\mathrm{TTI}}$ also increases, while the others reduce. For BW = 20MHz, this element is as large as $\Delta\mathcal{O}_{\mathrm{RS}}^{\mathrm{TTI}} = 91\%$. Moreover, $\Delta\mathcal{O}_{\mathrm{RS}}^{\mathrm{TTI}} + \Delta\mathcal{O}_{\mathrm{PDCCH}}^{\mathrm{TTI}}$ ranges between 84% and 97% in all the BW range 1.4 MHz $\leq$ BW $\leq$ 20 MHz. When applying the model at frame scale, the results exhibit the same trends, although the relative contributions of $\mathcal{O}_{\mathrm{RS}}^{\mathrm{frame}}$ and $\mathcal{O}_{\mathrm{PDCCH}}^{\mathrm{frame}}$ are slightly smaller. This is because, at frame scale, there are overhead components ($\mathcal{O}_{\mathrm{SCH}}^{\mathrm{frame}}$ and $\mathcal{O}_{\mathrm{PBCH}}^{\mathrm{frame}}$) that do *not* appear in a generic TTI, reducing thus the relative contribution of $\mathcal{O}_{\mathrm{RS}}^{\mathrm{frame}}$ and $\mathcal{O}_{\mathrm{PDCCH}}^{\mathrm{frame}}$, which, in any case, are clearly the dominant contributions to the total overhead.

The relative error made when computing the throughput at TTI level ($\mathcal{T}_{\mathrm{TOT}}^{\mathrm{TTI}}$) instead of at frame level ($\mathcal{T}_{\mathrm{TOT}}^{\mathrm{frame}}$), $\Delta\mathcal{T}_{\mathrm{TOT}}$, has been found to be $\Delta\mathcal{T}_{\mathrm{TOT}} < 1\%$ in those channel bandwidths for which LTE exhibits the highest data rates (BW

$= 10, 15, 20$ MHz). The practical consequence is that if an accuracy lower than 1% is admitted, there is *no* practical difference between computing the overhead either at TTI level or at frame level: $\mathcal{T}_{\mathrm{TOT}}^{\mathrm{frame}} \approx \mathcal{T}_{\mathrm{TOT}}^{\mathrm{TTI}} = \mathcal{T}$.

Finally, as all overhead mechanisms have been computed in detail, the real data throughput results in being slightly lower than those computed using less realistic models. In particular, the limiting value has been found to be $\mathcal{T} = 313.6$ Mbps.

2. Influence of the Hybrid ARQ configuration

The Physical Hybrid ARQ Indicator Channel (PHICH) carries the Hybrid-ARQ acknowledgment in DL (corresponding to its previous counterpart uplink data transfers). We have shown that its corresponding overhead $\mathcal{O}_{\mathrm{PHICH}}^{\mathrm{TTI}}$ depends, among others, on the PHICH Group Scaling Factor, N_g, which can only take values in the set $\{1/6, 1/2, 1, 2\}$. We have found that working with the case in which $N_g = 2$ (a worst case that maximizes $\mathcal{O}_{\mathrm{PHICH}}^{\mathrm{TTI}}$) only reduces the peak bit rate in $\approx 4\%$ when compared to that computed when using $N_g = 1/6$ (best case). This worst case leads to a bit rate 303 Mbps. If we consider as acceptable a maximum error of 4%, this will provide us a a safety factor that includes the possibility of using any allowed value of N_g.

3. Influence of MAT and resource allocation type

Multiple antennas technologies, which allows for increasing throughput, depend on the number of transmitting antennae (n), the number of layers (M), and the number of transmitted codewords (q). The resource allocation types defined in LTE lead to a number of allowed feasible combinations of values for n, N and q. Specifically, we have computed the peak data $\mathcal{T}$ (Mbps) as a function of BW (MHz), parametrized by n and the allowed values for q in five possible cases: 1) SISO (Single-Input and Single-Output, $n = 1$), with n =1 and $q = 0$ (only 1 codeword is allowed); 2) 2×2 MIMO (Multiple-Input and Multiple-Output) and $q = 0$; 3) 2×2 MIMO and $q = 1$; 4) 4×4 MIMO and $q = 0$; and 5) 4×4 MIMO and $q = 1$.

The first interesting result is that increasing the number of antennas n in MIMO schemes *not always* leads to a higher throughput. When using 2 codewords ($q = 1$), increasing n does lead to a superior performance than that of SISO for any BW value. However, when using only 1 codeword ($q = 0$), increasing n does *not* increase throughput but decreases it. For instance, for BW = 20 MHz, the peak data rate for SISO is $\mathcal{T}_{\mathrm{SISO}} \approx 85$ Mbps, which is greater than the one of 2×2 MIMO with $q = 0$ ($\mathcal{T}_{2\times 2\ \mathrm{MIMO}} \approx 82$ Mbps), and 4×4 MIMO with $q = 0$ ($\mathcal{T}_{4\times 4\ \mathrm{MIMO}} \approx 77$ Mbps). This is because of the additional reference signal resources required for higher order transmission schemes, which increase the percentage of overhead RE with respect to the available ones.

The second interesting result is related to the relative weight of each overhead component (with respect to the total) as a function of $n = 1, 2, 4$. In the three

transmitting configurations, for BW $\geq$ 5 MHz, the dominant overhead components are, at great extent, $\mathcal{O}_{\mathrm{RS}}$ and, to a lesser extent, $\mathcal{O}_{\mathrm{PHICH}}$. However, at BW = 1.4 MHz, as n reduces, the influence of $\mathcal{O}_{\mathrm{PDCCH}}$ increases in the detriment of $\mathcal{O}_{\mathrm{RS}}$: for $n = 4$, $\Delta\mathcal{O}_{\mathrm{RS}} \approx 52\%$ and $\Delta\mathcal{O}_{\mathrm{PDCCH}} \approx 25\%$, while, for $n = 1$, $\mathcal{O}_{\mathrm{PDCCH}}$ becomes clearly dominant ($\Delta\mathcal{O}_{\mathrm{PDCCH}} \approx 38\%$) when compared to the others.

4. Maximum number of users per TTI

The model includes the existence of a maximum number of users per TTI based on the fact that there is a limited number of controlling OFDM symbols for shared-data transmitting subframes. Specifically, we have computed the maximum number of users per TTI, $N_{\mathrm{UE}}^{\mathrm{TTI}_i}|_{\mathrm{MAX}}$, parametrized by $N_g = 1/6, 1/2, 1, 2$, as a function of BW, for different configurations: SISO with $q = 0$; 2×2 MIMO with $q = 0$ and $q = 1$; and $4{\times}4$ MIMO with $q = 0$ and $q = 1$. The main results are: 1) Increasing the number of antennae n reduces the maximum number of users per TTI, $N_{\mathrm{UE}}^{\mathrm{TTI}_i}|_{\mathrm{MAX}}$, because it requires higher $\mathcal{O}_{\mathrm{RS}}^{\mathrm{TTI}}$. 2) Increasing either N_g or q reduces $N_{\mathrm{UE}}^{\mathrm{TTI}_i}|_{\mathrm{MAX}}$ because $\mathcal{O}_{\mathrm{PHICH}}^{\mathrm{TTI}}$ increases.

4.2 Future work

The versatility and flexibility of the model presented in this thesis for performance evaluation of mobile broadband technologies makes it change sensitive. Any change in the configuration of the system is clearly reflected in the performance output (i.e. capacity and goodput). This allows the utilization of this model for detailed studies in which a set of different configurations can be simulated and compared to reach an optimum system performance, at each given deployment scenario. The results of these detailed studies can be programed in a visualization tool to assist the network planning and dimensioning tasks with an optimized design.

In addition to the attributes of dynamic overheads studied in this thesis, the impact of handover and mobility overheads shall be considered for enhanced estimation of system performance in future studies.

The service profile model presented in this thesis draws the foundations to arrive at a more elaborated model, which allows matching with consanguineously changing user's applications and habits. A real-time survey of the service profile, based on in-filed monitoring of the timely usage (hourly, daily, etc) of the system will allow adopting more practical values for the application distribution and contention ratios. The results of this future study can potentially replace the complexity of a real-time scheduling simulation with a simple and adoptable models for: best case, worse case and average case user serving scenarios, that can satisfy an accurate network planning and cost estimation purposes.

Finally, the proposed algorithm in this thesis for permanence evaluation of LTE [176] and WiMAX is global and system independent. The same approach

can be used to study and model the performance for latest cellular systems such as LTE-Advanced [177] and other technologies and enhancements that together build the 5G networks [178]. In particular, as pointed out in [179], the infrastructure densification that the ultra-dense networks (UDN) paradigm [180] requires (in which user density and node density can become similar [181]) can be put into practice by increasing the density of operator-deployed infrastructure nodes along with user-deployed access nodes and mobile user devices working as infrastructure prosumers. Such network densification (up to the point that the densities of serving nodes and served nodes become similar) along with massive MIMO [179, 182] are expected to become the cornerstones of 5G networks [178, 179], increasing capacity in several orders of magnitude. The technical details for overhead mechanisms of each technologies shall be adopted to be used as an input to the first process of the algorithm. Then using the same proposed incremental algorithm for capacity estimation, the service provision capabilities of each hot-spot can be evaluated.

List of Publications

1. **A.M. Ahmadzadeh**, J.E. Sanchez-García, B. Saavedra-Moreno, and A. Portilla-Figueras and S. Salcedo-Sanz, "Capacity estimation algorithm for simultaneous support of multi-class traffic services in Mobile WiMAX". *Computer Communications*, vol.35, no.1, pp. 109–119, 2012

2. Salcedo-Sanz, S., Sanchez-García, J. E., Portilla-Figueras, J. A., Jiménez-Fernández, S., and **Ahmadzadeh, A. M**. (2014). "A coral-reef optimization algorithm for the optimal service distribution problem in mobile radio access networks". *Transactions on Emerging Telecommunications Technologies*, 25(11), 1057-1069.

3. **A.M. Ahmadzadeh**, Lucas Cuadra, Miguel A. del Arco-Vega, J. Antonio Portilla-Figueras, Sancho Salcedo-Sanz. (2016) "Influence of overhead on LTE downlink performance: A comprehensive model", Submitted to *Telecommunication Systems*. Currently is under review.

4. Sánchez-García, J. E., **Ahmadzadeh, A. M**., Jiménez-Fernández, S., Salcedo-Sanz, S., Portilla-Figueras, J. A. (2011). "Impact of the HSPDA-based mobile broadband access on the investment of the 3G access network". In Mobile Lightweight Wireless Systems (pp. 303-311). Springer Berlin Heidelberg.

5. J.E Sánchez-García and **A. M Ahmadzadeh** and B. Saavedra-Moreno and S. Salcedo-Sanz and A. Portilla-Figueras. "Strategic Methods for Radio Access Design in 2G/3G Networks". 3 International ICST Conference on Mobile Lightweight Wireless Systems (Mobilight '11), Spain 9-11 May 2011, 4307–4311

References

[1] Ericsson. Ericsson Mobility Report. On the pulse of the networked society. Technical Report 1, Ericsson, available on https://www.ericsson.com/res/docs/2016/ericsson-mobility-report-2016.pdf, 2016.

[2] GSMA. The Mobile Economy 2015. Technical Paper available on www.gsmamobileeconomy.com, 9 2015.

[3] Ericsson. Ericsson Mobility Report. On the pulse of the networked society. Technical Report 1, Ericsson, available on http://www.ericsson.com/res/docs/2015/ericsson-mobility-report-june-2015.pdf, 2015.

[4] Ericsson. Understanding the Networked Society. Technical Report 1, Ericsson, available on https://www.ericsson.com/assets/local/networked-society/reports/wp-understanding-the-networked-society.pdf, 2015.

[5] Michael Miller. *The Internet of things: How smart TVs, smart cars, smart homes, and smart cities are changing the world*. Pearson Education, 2015.

[6] Shaibal Chakrabarty and Daniel W Engels. A secure IoT architecture for Smart Cities. In *2016 13th IEEE Annual Consumer Communications & Networking Conference (CCNC)*, pages 812–813. IEEE, 2016.

[7] Vito Albino, Umberto Berardi, and Rosa Maria Dangelico. Smart cities: Definitions, dimensions, performance, and initiatives. *Journal of Urban Technology*, 22(1):3–21, 2015.

[8] Ericsson. Stockholm tops Ericsson Networked Society City Index 2016. Technical Report 1, Ericsson, available on http://hugin.info/1061/R/2021689/751081.pdf, 2016.

[9] Ericsson. Laying the foundations for a smart, sustainable city. Technical Report 1, Ericsson, White paper Uen 284 23-3277, January 2016, available on https://www.ericsson.com/res/docs/whitepapers/wp-smart-cities.pdf, 2016.

[10] GSA Association. HSPA Operator Commitments report: 582 launched, incl. 182 42Mbs DC-HSPA+ and first 63Mbps 3C-HSPA+. Information Paper available on http://www.gsacom.com, 7 2015.

[11] Thomas Chapman, Erik Larsson, PETER von Wrycza, Erik Dahlman, Stefan Parkvall, and Johan Skold. *HSPA Evolution: The Fundamentals for Mobile Broadband*. Academic Press, 2014.

[12] Harri Holma and Antti Toskala. *WCDMA for UMTS: HSPA evolution and LTE*. John Wiley & Sons, 2010.

[13] Erik Dahlman, Stefan Parkvall, Johan Skold, and Per Beming. *3G evolution: HSPA and LTE for mobile broadband*. Academic press, 2010.

[14] Stefania Sesia, Issam Toufik, and Matthew Baker. *LTE: the UMTS long term evolution*. Wiley Online Library, 2009.

[15] Xiaoli Chu, David Lopez-Perez, Yang Yang, and Fredrik Gunnarsson. *Heterogeneous Cellular Networks: Theory, Simulation and Deployment*. Cambridge University Press, 2013.

[16] Jyrki TJ Penttinen. *The Telecommunications Handbook: Engineering Guidelines for Fixed, Mobile and Satellite Systems*. John Wiley & Sons, 2015.

[17] Jeffrey G Andrews, Holger Claussen, Mischa Dohler, Sundeep Rangan, and Mark C Reed. Femtocells: Past, present, and future. *Selected Areas in Communications, IEEE Journal on*, 30(3):497–508, 2012.

[18] Holger Claussen, Lester TW Ho, and Louis G Samuel. An overview of the femtocell concept. *Bell Labs Technical Journal*, 13(1):221–245, 2008.

[19] Vikram Chandrasekhar, Jeffrey G Andrews, and Alan Gatherer. Femtocell networks: a survey. *Communications Magazine, IEEE*, 46(9):59–67, 2008.

[20] Ashok Kumar and Peter Jarich. 3G Wireless Investment Motivations: Why Do Mobile Operators Continue to Invest in 3G Wireless Network Infrastructure? Technical Paper available on http://www.currentanalysis.com/, 9 2015.

[21] L Cuadra, A Aybar-Ruíz, MA del Arco, J Navío-Marco, JA Portilla-Figueras, and S Salcedo-Sanz. A Lamarckian Hybrid Grouping Genetic Algorithm with repair heuristics for resource assignment in WCDMA networks. *Applied Soft Computing*, 43:619–632, 2016.

[22] GSA. GSM/3G Stats. Fast Facts. Q1 2015 mobile subscriptions. Technical report, Global mobile Suppliers Association (GSA), available on http://www.gsacom.com/, 2015.

[23] Mahmoud AM Albreem and Nur Afiqah Hani Binti Ismail. A review: detection techniques for LTE system. *Telecommunication Systems*, pages 1–16, 2015.

[24] Tuan Ta and John S Baras. Exploiting diversity of usage to enhance user equipment energy efficiency in LTE networks. *Telecommunication Systems*, 59(1):5–23, 2015.

[25] Bart Sas, Elena Bernal-Mor, Kathleen Spaey, Vicent Pla, Chris Blondia, and Jorge Martinez-Bauset. Modelling the time-varying cell capacity in LTE networks. *Telecommunication Systems*, 55(2):299–313, 2014.

[26] Xi Li, Umar Toseef, Dominik Dulas, Wojciech Bigos, Carmelita Görg, Andreas Timm-Giel, and Andreas Klug. Dimensioning of the LTE access network. *Telecommunication Systems*, 52(4):2637–2654, 2013.

[27] Desta Haileselassie Hagos. The performance of network-controlled mobile data offloading from LTE to WiFi networks. *Telecommunication Systems*, pages 1–20, 2015.

[28] Adyson M Maia, Dario Vieira, Miguel F de Castro, and Yacine Ghamri-Doudane. A fair QoS-aware dynamic LTE scheduler for machine-to-machine communication. *Computer Communications*, 2016.

[29] Sasan Adibi, Amin Mobasher, and Tom Tofigh. LTE networking: extending the reach for sensors in mHealth applications. *Transactions on Emerging Telecommunications Technologies*, 25(7):692–706, 2014.

[30] Jason Brown and Jamil Y Khan. Key performance aspects of an LTE FDD based smart grid communications network. *Computer Communications*, 36(5):551–561, 2013.

[31] Mustafa Ismael Salman, Chee Kyun Ng, Nor Kamariah Noordin, Borhanuddin Mohd Ali, and Aduwati Sali. A self-configured link adaptation for green LTE downlink transmission. *Transactions on Emerging Telecommunications Technologies*, 26(2):258–275, 2015.

[32] Ayad Atiyah Abdulkafi, David Chieng, Tiong Sieh Kiong, Alvin Ting, Johnny Koh, and Abdulaziz M Ghaleb. Energy-aware load adaptive framework for LTE heterogeneous network. *Transactions on Emerging Telecommunications Technologies*, 25(9):943–953, 2014.

[33] Liljana Gavrilovska, Jaap van de Beek, Yong Xie, Erik Lidström, Janne Riihijärvi, Petri Mähönen, Vladimir Atanasovski, Daniel Denkovski, and Valentin Rakovic. Enabling LTE in TVWS with radio environment maps: From an architecture design towards a system level prototype. *Computer Communications*, 53:62–72, 2014.

[34] Christos Papathanasiou, Nikos Dimitriou, and Leandros Tassiulas. Dynamic radio resource and interference management for MIMO–OFDMA mobile broadband wireless access systems. *Computer Networks*, 57(1):3–16, 2013.

[35] Junyi Li, Xinzhou Wu, and Rajiv Laroia. *OFDMA mobile broadband communications: A systems approach*. Cambridge University Press, 2013.

[36] Christopher Cox and Baggy Cox. *An Introduction to LTE: LTE, LTE-Advanced, SAE, VoLTE and 4G Mobile Communications*. John Wiley & Sons, 2014.

[37] Erik Dahlman, Stefan Parkvall, and Johan Skold. *4G: LTE/LTE-advanced for mobile broadband*. Academic press, 2013.

[38] Maciej Stasiak, Mariusz Glabowski, Arkadiusz Wisniewski, and Piotr Zwierzykowski. *Modelling and dimensioning of mobile wireless networks: from GSM to LTE*. John Wiley & Sons, 2010.

[39] Ursula Challita, Zaher Dawy, George Turkiyyah, and Joe Naoum-Sawaya. A chance constrained approach for LTE cellular network planning under uncertainty. *Computer Communications*, 73:34–45, 2016.

[40] I. Siomina and Di Yuan. Analysis of Cell Load Coupling for LTE Network Planning and Optimization. *IEEE Transactions on Wireless Communications*, 11(6):2287–2297, 2012.

[41] Kurt Majewski and Michael Koonert. Conservative cell load approximation for radio networks with Shannon channels and its application to LTE network planning. In *Telecommunications (AICT), 2010 Sixth Advanced International Conference on*, pages 219–225. IEEE, 2010.

[42] Luca Anchora, Marco Mezzavilla, Leonardo Badia, and Michele Zorzi. A performance evaluation tool for spectrum sharing in multi-operator LTE networks. *Computer Communications*, 35(18):2218–2226, 2012.

[43] Kai Yang, Steven Martin, and Tara Ali Yahiya. LTE uplink interference aware resource allocation. *Computer Communications*, 66:45–53, 2015.

[44] Dionysis Xenakis, Nikos Passas, and Christos Verikoukis. An energy-centric handover decision algorithm for the integrated LTE macrocell-femtocell network. *Computer Communications*, 35(14):1684–1694, 2012.

[45] GSMA. How Mobile Investment Can Lead the World out of Financial Crisis. Technical Paper available on http://www.itu.int/osg/csd/emerging-trends/crisis/fc11.html, 6 2009.

[46] Phillippa Biggs and Stephane Rollet. *Confronting the Crisis: Its Impact on the ICT Industry*. International Telecommunication Union, Corporate Strategy Division, 2009.

[47] The Mobile World. The impact of the current financial crisis on the telecom industry. Technical Paper available on http://www.itu.int/osg/csd/emerging-trends/crisis/fc11.html, 6 2009.

[48] O Gautam, VK Singh, and R Sharma. Global Recession and Its Impact on Telecommunication Industry: An Empirical Dissection. *International Journal of Management and Business Research*, 5(2):107–116, 2015.

[49] Joss Gillet. *Mobile operators rein in costs to ride-out recession; prepare to ramp-up network investments* (Wireless Intelligence). Technical report, GSMA Intelligence, 9 2011.

[50] CMT. *Global Annual Report* (http://cmtdata.cmt.es/). Technical report, Spanish National Regulatory Authority (CMT, Comisión del Mercado de las Telecomunicaciones), 11 2012.

[51] Sancho Salcedo-Sanz, Juan Eulogio Sanchez-Garcia, José Antonio Portilla-Figueras, Silvia Jimenez-Fernandez, and Amir M Ahmadzadeh. A coral-reef optimization algorithm for the optimal service distribution problem in mobile radio access networks. *Transactions on Emerging Telecommunications Technologies*, 25(11):1057–1069, 2014.

[52] H Vicky Zhao and Weifeng Su. Cooperative wireless multicast: performance analysis and power/location optimization. *Wireless Communications, IEEE Transactions on*, 9(6):2088–2100, 2010.

[53] Yiqing Zhou, Hang Liu, Zhengang Pan, Lin Tian, Jinglin Shi, and Guanghua Yang. Two-stage cooperative multicast transmission with optimized power consumption and guaranteed coverage. *Selected Areas in Communications, IEEE Journal on*, 32(2):274–284, 2014.

[54] Bo Li, Yang Qin, Chor Ping Low, and Choon Lim Gwee. A survey on mobile WiMAX [wireless broadband access]. *Communications Magazine, IEEE*, 45(12):70–75, 2007.

[55] Sassan Ahmadi. An overview of next-generation mobile WiMAX technology. *Communications Magazine, IEEE*, 47(6):84–98, 2009.

[56] Morris J Chang, Zakhia Abichar, and Chau-Yun Hsu. WiMAX or LTE: Who will lead the broadband mobile Internet? *IT Professional Magazine*, 12(3):26, 2010.

[57] S Srikanth, P Pandian, and Xavier Fernando. Orthogonal frequency division multiple access in WiMAX and LTE: a comparison. *Communications Magazine, IEEE*, 50(9):153–161, 2012.

[58] Ronny Yongho Kim, Inuk Jung, Xiangying Yang, and Chao-Chin Chou. Advanced handover schemes in IMT-advanced systems [WiMAX/LTE Update]. *Communications Magazine, IEEE*, 48(8):78–85, 2010.

[59] Nageen Himayat, Shilpa Talwar, Anil Rao, and Robert Soni. Interference management for 4G cellular standards [WiMAX/LTE update]. *Communications Magazine, IEEE*, 48(8):86–92, 2010.

[60] Qinghua Li, Guangjie Li, Wookbong Lee, Moon-il Lee, David Mazzarese, Bruno Clerckx, and Zexian Li. MIMO techniques in WiMAX and LTE: a feature overview. *Communications Magazine, IEEE*, 48(5):86–92, 2010.

[61] I-Kang Fu, Yih-Shen Chen, Paul Cheng, Youngsoo Yuk, Ronny Yongho Kim, Jin Sam Kwak, et al. Multicarrier technology for 4G WiMAX system [WiMAX/LTE update]. *Communications Magazine, IEEE*, 48(8):50–58, 2010.

[62] Mehdi Alasti, Behnam Neekzad, Jie Hui, and Rath Vannithamby. Quality of service in WiMAX and LTE networks. *IEEE Communications Magazine*, 48(5):104–111, 2010.

[63] Ozgur Oyman, Jeffrey Foerster, Yong-joo Tcha, and Seong-Choon Lee. Toward enhanced mobile video services over WiMAX and LTE [WiMAX/LTE update]. *Communications Magazine, IEEE*, 48(8):68–76, 2010.

[64] Sinan Ghassan Abid Ali, Mohd Dani Baba, Mohd Asri Mansor, and Labeeb Mohsin Abdullah. An IMS signalling module for LTE-based femtocell networks. In *Control and System Graduate Research Colloquium (ICSGRC), 2014 IEEE 5th*, pages 247–252. IEEE, 2014.

[65] Sinan Ghassan Abid Ali, Mohd Dani Baba, Mohd Asri Mansor, and Labeeb Mohsin Abdullah. SIP based IMS registration signalling for LTE-based femtocell networks. In *Control and System Graduate Research Colloquium (ICSGRC), 2014 IEEE 5th*, pages 25–30. IEEE, 2014.

[66] M Rinne and Olav Tirkkonen. LTE, the radio technology path towards 4G. *Computer Communications*, 33(16):1894–1906, 2010.

[67] Fahimeh Rezaei, Michael Hempel, and Hamid Sharif. LTE PHY performance analysis under 3GPP standards parameters. In *Computer Aided Modeling and Design of Communication Links and Networks (CAMAD), 2011 IEEE 16th International Workshop on*, pages 102–106. IEEE, 2011.

[68] IEEE. IEEE Standard 802.16e-2005, IEEE standard for local and metropolitan area networks part 16: air interface for fixed and mobile broadband wireless access system. Proceedings of IEEE Std., 2 2006.

[69] A.M. Ahmadzadeh, J.E. Sanchez-García, B. Saavedra-Moreno, A. Portilla-Figueras, and S. Salcedo-Sanz. Capacity estimation algorithm for simultaneous support of multi-class traffic services in Mobile WiMAX. *Computer Communications*, 35(1):109–119, 2012.

[70] Sassan Ahmadi. *Mobile WiMAX: A systems approach to understanding IEEE 802.16 m radio access technology.* Academic Press, 2010.

[71] Iana Siomina, Anders Furuskar, and Gabor Fodor. A mathematical framework for statistical QoS and capacity studies in OFDM networks. In *Personal, Indoor and Mobile Radio Communications, 2009 IEEE 20th International Symposium on*, pages 2772–2776. IEEE, 2009.

[72] Dongzhe Cui. LTE peak rates analysis. In *Wireless and Optical Communications Conference, 2009. WOCC 2009. 18th Annual*, pages 1–3. IEEE, 2009.

[73] Loutfi Nuaymi and Ziad Noun. Simple Capacity Estimation in WiMAX/802.16 Systems. In *Personal, Indoor and Mobile Radio Communications, 2006 IEEE 17th International Symposium on*, pages 1–5. IEEE, 2006.

[74] Bong-Ho Kim, Jungnam Yun, Yerang Hur, Chakchai So-In, Raj Jain, and A-K Al Tamimi. Capacity estimation and TCP performance enhancement over mobile WiMAX networks. *Communications Magazine, IEEE*, 47(6):132–141, 2009.

[75] Sergey N Moiseev and Mikhail S Kondakov. Fast system load estimation in the IEEE 802.16 OFDMA network. *Computer Communications*, 32(7):1228–1232, 2009.

[76] Stanislav A Filin, Sergey N Moiseev, Mikhail S Kondakov, Alexandre V Garmonov, Jaeho Lee, et al. QoS-Guaranteed Cross-Layer Adaptive Transmission Algorithms for the IEEE 802.16 OFDMA Systems. In *Wireless Communications and Networking Conference, 2006. WCNC 2006. IEEE*, volume 2, pages 964–971. IEEE, 2006.

[77] Jiho Jang and Kwang Bok Lee. Transmit power adaptation for multiuser OFDM systems. *IEEE Journal on Selected Areas in Communications*, 21(2):171–178, 2003.

[78] Sławomir Stańczak, Angela Feistel, Marcin Wiczanowski, and Holger Boche. Utility-based power control with QoS support. *Wireless Networks*, 16(6):1691–1705, 2010.

[79] Marcin Wiczanowski, Holger Boche, and Slawomir Stanczak. An algorithm for optimal resource allocation in cellular networks with elastic traffic. *IEEE Transactions on Communications*, 57(1):41–44, 2009.

[80] Marcin Wiczanowski, Sławomir Stańczak, and Holger Boche. Providing quadratic convergence of decentralized power control in wireless networks – The method of min-max functions. *Signal Processing, IEEE Transactions on*, 56(8):4053–4068, 2008.

[81] Sławomir Stańczak, Marcin Wiczanowski, and Holger Boche. Distributed utility-based power control: objectives and algorithms. *Signal Processing, IEEE Transactions on*, 55(10):5058–5068, 2007.

[82] D. Astély, E. Dahlman, A. Furuskar, Y. Jading, M. Lindström, and S. Parkvall. LTE: the evolution of mobile broadband. *IEEE Communications Magazine*, 47(4):44–51, 2009.

[83] Anna Larmo, Magnus Lindstrom, Michael Meyer, Ghyslain Pelletier, Johan Torsner, and Henning Wiemann. The LTE link-layer design. *Communications Magazine, IEEE*, 47(4):52–59, 2009.

[84] Ziqi Zhang, Zhuyan Zhao, Hao Guan, Deshan Miao, and Zhenhui Tan. Study of signaling overhead caused by keep-alive messages in LTE network. In *Vehicular Technology Conference (VTC Fall), 2013 IEEE 78th*, pages 1–5. IEEE, 2013.

[85] Seyed Mohsen Razavi and Di Yuan. Reducing signaling overhead by overlapping tracking area list in LTE. In *Wireless and Mobile Networking Conference (WMNC), 2014 7th IFIP*, pages 1–7. IEEE, 2014.

[86] Sara Modarres Razavi, Di Yuan, Fredrik Gunnarsson, and Johan Moe. Exploiting tracking area list for improving signaling overhead in LTE. In *Vehicular Technology Conference (VTC 2010-Spring), 2010 IEEE 71st*, pages 1–5. IEEE, 2010.

[87] Lamia Osman Widaa Osman. *Optimization of Tracking Area List Design for Reducing Total Signaling Overhead in LTE Systems*. PhD thesis, University of Khartoum, 2015.

[88] Payal Mittal, Izzat Darwazeh, and Hayk Manukyan. Overhead estimation during intra eNB handover in 4G LTE systems. In *Communication Systems, Networks & Digital Signal Processing (CSNDSP), 2014 9th International Symposium on*, pages 960–965. IEEE, 2014.

[89] Ying Yang, Pengfei Li, and Weidong Wang. Algorithm about mobility load balance considering system overhead on LTE system. In *Wireless and Optical Communication Conference (WOCC), 2013 22nd*, pages 231–235. IEEE, 2013.

[90] Muhammad Basit Shahab. Efficient channel quality indicator reporting schemes in LTE with reduced signaling overhead. In *Telecommunications and Signal Processing (TSP), 2015 38th International Conference on*, pages 123–128. IEEE, 2015.

[91] Joseph Stalin Thainesh, Ning Wang, and Rahim Tafazolli. Reduction of core network signalling overhead in cluster based LTE small cell networks. In *Computer Aided Modelling and Design of Communication Links and Networks (CAMAD), 2015 IEEE 20th International Workshop on*, pages 226–230. IEEE, 2015.

[92] Petter Edström. Overhead Impacts on Long-Term Evolution Radio Networks. Master's thesis, KTH Information and Communication Technology, Stockholm, 2007.

[93] Jun-Bae Seo and Victor CM Leung. Performance modeling and stability of semi-persistent scheduling with initial random access in LTE. *Wireless Communications, IEEE Transactions on*, 11(12):4446–4456, 2012.

[94] Ajay Pratap and Hemanta Kumar Pati. Capacity Estimation for Cellular LTE Using AMR Codec with Semi-persistent Scheduling. In *Intelligent Computing, Communication and Devices*, pages 725–736. Springer, 2015.

[95] Lu Ding, Fei Tong, Zhen Chen, and Zhimin Liu. A novel MCS selection criterion for VoIP in LTE. In *Wireless Communications, Networking and Mobile Computing (WiCOM), 2011 7th International Conference on*, pages 1–4. IEEE, 2011.

[96] Dajie Jiang, Haiming Wang, Esa Malkamaki, and Esa Tuomaala. Principle and performance of semi-persistent scheduling for VoIP in LTE system. In *Wireless Communications, Networking and Mobile Computing, 2007. WiCom 2007. International Conference on*, pages 2861–2864. IEEE, 2007.

[97] Afif Osseiran, Jose F Monserrat, and Werner Mohr. *Mobile and Wireless Communications for IMT-advanced and Beyond*. John Wiley & Sons, 2011.

[98] Moray Rumney. IMT-Advanced: 4G Wireless Takes Shape in an Olympic Year Agilent Measurement Journal, September 1, 2008 IEEE 802.16 TGm.

[99] IEEE. *IEEE Standard for Local and Metropolitan Area networks Part 16: Air Interface for Fixed Broadband Wireless Access Systems* (IEEE Std 802.16-2004). Technical report, IEEE, 10 2004.

[100] IEEE. *DRAFT Standard for Local and Metropolitan Area Networks, Part 16: Air Interface for Broadband Wireless Access Systems* (IEEE P802.16Rev2/D9a). Technical report, IEEE, 3 2009.

[101] WiMAX Forum. *Mobile System Profile Specification, Release 1.5 - Common Part* (WiMAX Forum Specification (2009-08-01)). Technical report, WiMAX Forum, 8 2009.

[102] Jeffrey G Andrews, Arunabha Ghosh, and Rias Muhamed. *Fundamentals of WiMAX: Understanding Broadband Wireless Networking*. Prentice Hall, Pearson Education, 2007.

[103] Yang Xiao. *WiMAX/MobileFi: Advanced Research and Technology*. Auerbach Publications, Taylor & Francis Group, CRC Press, 2008.

[104] Raj Jain, Chakchai So-In, and Abdel-karim Al Tamimi. System-level Modeling of IEEE 802.16e - Mobile WiMAX Networks: Key Issues. *Wireless Communications, IEEE*, 15(5):73–79, 2008.

[105] Tara Ali-Yahiya, André-Luc Beylot, and Guy Pujolle. An adaptive cross-layer design for multiservice scheduling in OFDMA based mobile WiMAX systems. *Computer Communications*, 32(3):531–539, 2009.

[106] Y Ahmet Şekercioğlu, Milosh Ivanovich, and Alper Yeğin. A survey of MAC based QoS implementations for WiMAX networks. *Computer Networks*, 53(14):2517–2536, 2009.

[107] Carsten Ball, Thomas Hindelang, Iavor Kambourov, and Sven Eder. Spectral efficiency assessment and radio performance comparison between LTE and WiMAX. In *Personal, Indoor and Mobile Radio Communications, 2008. PIMRC 2008. IEEE 19th International Symposium on*, pages 1–6. IEEE, 2008.

[108] Zakhia Abichar, Yanlin Peng, and J Morris Chang. WiMAX: The emergence of wireless broadband. *IT professional*, 8(4):44–48, 2006.

[109] Koon Hoo Teo, Zhifeng Tao, and Jinyun Zhang. The mobile broadband WiMAX standard [standards in a nutshell]. *Signal Processing Magazine, IEEE*, 24(5):144–148, 2007.

[110] Yang Yang, Honglin Hu, Jing Xu, and Guoqiang Mao. Relay technologies for WiMAX and LTE-advanced mobile systems. *Communications Magazine, IEEE*, 47(10):100–105, 2009.

[111] Sassan Ahmadi. An Overview of Next Generation Mobile WiMAX: Technology and Prospects. *WiMAX Evolution*, page 441, 2009.

[112] Won-Hyoung Park, Sunghyun Cho, and Saewoong Bahk. Scheduler design for multiple traffic classes in OFDMA networks. *Computer Communications*, 31(1):174–184, 2008.

[113] Fellah, A. *WiMAX Market Trends and Deployments*, WiMAXCounts Newsletter. Technical report, Wimax Wifi Technology & Broadband Internet, available on http://www.wimaxcounts.com/, 2007.

[114] Yehuda Ben-Shimol, Itzik Kitroser, and Yefim Dinitz. Two-Dimensional Mapping for Wireless OFDMA Systems. *Broadcasting, IEEE Transactions on*, 52(3):388–396, 2006.

[115] Takeo Ohseki, Megumi Morita, and Takashi Inoue. Burst Construction and Packet Mapping Scheme for OFDMA Downlinks in IEEE 802.16 systems. In *Global Telecommunications Conference, 2007. GLOBECOM'07. IEEE*, pages 4307–4311. IEEE, 2007.

[116] 3GPP. *Physical Layer Aspects for Evolved UTRA* (3GPP Std 25.814 V1.2.3) . Technical report, 3GPP, 5 2006.

[117] Jianfeng Chen, Wenhua Jiao, and Hongxi Wang. A Service Flow Management Strategy for IEEE 802.16 Broadband Wireless Access Systems in TDD Mode. In *Communications, 2005. ICC 2005. 2005 IEEE International Conference on*, volume 5, pages 3422–3426. IEEE, 2005.

[118] Joswill Victor Pajaro Rodriguez and Amir Esmailpour. Integrated QoS provisioning for unified LTE-WiMAX networks. In *2016 International Conference on Computing, Networking and Communications (ICNC)*, pages 1–6. IEEE, 2016.

[119] Maryam Roodaki, Kaamran Raahemifar, and Bijan Raahemi. Analysis of Quality of Services in LTE and Mobile WiMAX. *Computers & Electrical Engineering*, 40(5):1508–1523, 2014.

[120] Sungjoo Lee, Chanwoo Cho, Een-kee Hong, and Byungun Yoon. Forecasting mobile broadband traffic: Application of scenario analysis and Delphi method. *Expert Systems with Applications*, 44:126–137, 2016.

[121] Tuan-Che Chen, Ying-Yu Chen, and Jyh-Cheng Chen. An efficient energy saving mechanism for IEEE 802.16 e wireless MANs. *Wireless Communications, IEEE Transactions on*, 7(10):3708–3712, 2008.

[122] Tuan-Che Chen, Jyh-Cheng Chen, and Ying-Yu Chen. Maximizing unavailability interval for energy saving in IEEE 802.16 e wireless MANs. *Mobile Computing, IEEE Transactions on*, 8(4):475–487, 2009.

[123] Yu-Chee Tseng, Jen-Jee Chen, and Yen-Chih Yang. Managing power saving classes in IEEE 802.16 wireless MANs: a fold-and-demultiplex method. *Mobile Computing, IEEE Transactions on*, 10(9):1237–1247, 2011.

[124] Jenhui Chen and Yu-Lin Li. Adaptive traffic indication algorithm for energy efficiency in IEEE 802.16 e systems. In *Communications and Networking in China (CHINACOM), 2010 5th International ICST Conference on*, pages 1–5. IEEE, 2010.

[125] Wen-Hwa Liao, Ssu-Chi Kuai, and Chen Liu. Efficient scheduling algorithm for multiple mobile subscriber stations with unsolicited grant service in an IEEE 802.16 e network. *Journal of Network and Computer Applications*, 63:50–55, 2016.

[126] Wen-Hwa Liao, Kuei-Ping Shih, and Nien-Tsung Lee. Integrated power-saving scheduling algorithm in IEEE 802.16 e networks. *Communications, IET*, 7(3):255–262, 2013.

[127] Wen-Hwa Liao, Sital Prasad Kedia, and Avinash Kumar Dubey. Scheduling and channel assignment algorithm for IEEE 802.16 mesh networks using clique partitioning technique. *Computer Communications*, 35(16):2025–2034, 2012.

[128] Shih-Chang Huang, Chien Chen, Rong-Hong Jan, and Cheng-Chung Hsieh. An energy-efficient scheduling for multiple MSSs in IEEE 802.16 e broadband wireless. In *Personal, Indoor and Mobile Radio Communications, 2008. PIMRC 2008. IEEE 19th International Symposium on*, pages 1–5. IEEE, 2008.

[129] Shih-Chang Huang, Rong-Hong Jan, and Chien Chen. Energy efficient scheduling with qos guarantee for ieee 802.16 e broadband wireless access networks. In *Proceedings of the 2007 international conference on Wireless communications and mobile computing*, pages 547–552. ACM, 2007.

[130] Jinglin Shi, Gengfa Fang, Yi Sun, Jihua Zhou, Zhongcheng Li, and Eryk Dutkiewicz. WLC17-5: improving mobile station energy efficiency in IEEE 802.16 e WMAN by burst scheduling. In *Global Telecommunications Conference, 2006. GLOBECOM'06. IEEE*, pages 1–5. IEEE, 2006.

[131] Chia-Yen Lin, Hsi-Lu Chao, Yang-Jang Liao, and Tzu-Jane Tsai. Least-Awake-Frame Scheduling with delay guarantee for IEEE 802.16 e broadband wireless access networks. In *Personal, Indoor and Mobile Radio Communications, 2008. PIMRC 2008. IEEE 19th International Symposium on*, pages 1–5. IEEE, 2008.

[132] Chia-Yen Lin and Hsi-Lu Chao. Energy-saving scheduling in IEEE 802.16 e networks. In *Local Computer Networks, 2008. LCN 2008. 33rd IEEE Conference on*, pages 130–135. IEEE, 2008.

[133] Chakchai So-In, Raj Jain, and Abdel-Karim Tamimi. Scheduling in IEEE 802.16 e mobile WiMAX networks: key issues and a survey. *Selected Areas in Communications, IEEE Journal on*, 27(2):156–171, 2009.

[134] Shiann-Tsong Sheu, Yen-Chieh Cheng, Lu-Wei Chen, Jung-Shyr Wu, and Johnson Chang. Listening interval spreading approach (lisa) for handling burst traffic in ieee 802.16 e wireless metropolitan area networks. In *Vehicular Technology Conference, 2008. VTC 2008-Fall. IEEE 68th*, pages 1–5. IEEE, 2008.

[135] Umang R Mori, Parth M Chandarana, Gunjan V Gajjar, and Shivakrishna Dasi. Performance comparison of different modulation schemes in advanced technologies WiMAX and LTE. In *Advance Computing Conference (IACC), 2015 IEEE International*, pages 286–289. IEEE, 2015.

[136] Junaid Ahmed Zubairi, Erdem Erdogan, and Shaun Reich. Experiments in fair scheduling in 4G WiMAX and LTE. In *High Performance Computing & Simulation (HPCS), 2015 International Conference on*, pages 277–282. IEEE, 2015.

[137] Leo Yi, Kai Miao, and Adrian Liu. A comparative study of WiMAX and LTE as the next generation mobile enterprise network. In *Advanced Communication Technology (ICACT), 2011 13th International Conference on*, pages 654–658. IEEE, 2011.

[138] Vasile Horia Muntean and Marius Otesteanu. WiMAX versus LTE-An overview of technical aspects for next generation networks technologies. In *Electronics and Telecommunications (ISETC), 2010 9th International Symposium on*, pages 225–228. IEEE, 2010.

[139] Margot Deruyck, Willem Vereecken, Emmeric Tanghe, Wout Joseph, Mario Pickavet, Luc Martens, and Piet Demeester. Comparison of power consumption of mobile WiMAX, HSPA and LTE access networks. In *Telecommuni-*

cations Internet and Media Techno Economics (CTTE), 2010 9th Conference on, pages 1–7. IEEE, 2010.

[140] E Prince Edward. A Novel Seamless Handover Scheme for WiMAX/LTE Heterogeneous Networks. *Arabian Journal for Science and Engineering*, pages 1–15, 2015.

[141] Nadine Akkari. An IMS-based integration architecture for WiMax/LTE handover. *Computer Networks*, 57(18):3790–3798, 2013.

[142] Tara A Yahiya and Hakima Chaouchi. On the integration of LTE and mobile WiMAX networks. In *Computer Communications and Networks (ICCCN), 2010 Proceedings of 19th International Conference on*, pages 1–5. IEEE, 2010.

[143] Mustafa Ergen. *Mobile broadband: including WiMAX and LTE*. Springer Science & Business Media, 2009.

[144] Abd-Elhamid M Taha, Najah Abu Ali, and Hossam S Hassanein. *LTE, LTE-advanced and WiMAX: Towards IMT-advanced Networks*. John Wiley & Sons, 2011.

[145] Henrik Martikainen, Olli Alanen, and Alexander Sayenko. ARQ parameters for VoIP in IEEE 802.16 networks. In *Wireless Telecommunications Symposium, 2009. WTS 2009*, pages 1–6. IEEE, 2009.

[146] Loutfi Nuaymi. *WiMAX: technology for broadband wireless access*. John Wiley & Sons, 2007.

[147] WiMAX Forum. *Mobile WiMAX - Part I: A Technical Overview and Performance Evaluation* . Technical report, WiMAX Forum, 8 2006.

[148] 3GPP. LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Multiplexing and channel coding. Technical report, ETSI, France, 2010. Technical Specification Name: 3GPP TS 36.212 version 9.2.0, Release 9.

[149] 3GPP. LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures. Technical report, ETSI, France, 2010. Technical Specification Name: 3GPP TS 36.213 version 9.2.0 Release 9.

[150] Juho Lee, Jin-Kyu Han, and Jianzhong Charlie Zhang. MIMO technologies in 3GPP LTE and LTE-advanced. *EURASIP Journal on Wireless Communications and Networking*, 2009(1):1–10, 2009.

[151] Jonathan Duplicy, Biljana Badic, Rajarajan Balraj, Rizwan Ghaffar, Péter Horváth, Florian Kaltenberger, Raymond Knopp, István Z Kovács, Hung T Nguyen, Deepaknath Tandur, et al. MU-MIMO in LTE Systems. *EURASIP Journal on Wireless Communications and Networking*, 2011(1):1–13, 2011.

[152] Carsten F Ball, Robert Mullner, Johann Lienhart, and Hubert Winkler. Performance analysis of Closed and Open loop MIMO in LTE. In *Wireless Conference, 2009. EW 2009. European*, pages 260–265. IEEE, 2009.

[153] Francesco Capozzi, Giuseppe Piro, Luigi Alfredo Grieco, Gennaro Boggia, and Pietro Camarda. Downlink packet scheduling in LTE cellular networks: Key design issues and a survey. *Communications Surveys & Tutorials, IEEE*, 15(2):678–700, 2013.

[154] Stefan Schwarz, Christian Mehlfuhrer, and Markus Rupp. Low complexity approximate maximum throughput scheduling for LTE. In *Signals, Systems and Computers (ASILOMAR), 2010 Conference Record of the Forty Fourth Asilomar Conference on*, pages 1563–1569. IEEE, 2010.

[155] Raymond Kwan, Cyril Leung, and Jie Zhang. Proportional fair multiuser scheduling in LTE. *Signal Processing Letters, IEEE*, 16(6):461–464, 2009.

[156] Anders Furuskar, Tomas Jonsson, and Magnus Lundevall. The LTE radio interface-key characteristics and performance. In *Personal, Indoor and Mobile Radio Communications, 2008. PIMRC 2008. IEEE 19th International Symposium on*, pages 1–5. IEEE, 2008.

[157] 3GPP. LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Base Station (BS) radio transmission and reception. Technical report, ETSI, France, 2011. Technical Specification Name: 3GPP TS 36.104 version 10.2.0 Release 10.

[158] 3GPP. LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures (3GPP TS 36.213 version 10.10.0 Release 10). Technical report, 3GPP, 7 2013. ETSI TS 136 213 V10.10.0 (2013-07).

[159] 3GPP. LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical channels and modulation. Technical report, ETSI, France, 2010. Technical Specification Name: 3GPP TS 36.211 version 9.1.0, Release 9.

[160] Real-Wireless. 4G Capacity Gains v1.4. Technical report, OFCOM, 1 2011.

[161] Anirban Bhowal, Tanay Paradkar, and Neetesh Purohit. Evaluation of extended CP and ultra-extended CP in MBSFN. In *2015 IEEE UP Section Conference on Electrical Computer and Electronics (UPCON)*, pages 1–6. IEEE, 2015.

[162] 3GPP. LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical channels and modulation (3GPP TS 36.211 version 10.0.0 Release 10). Technical report, 3GPP, 1 2011. ETSI TS 136 211 V10.0.0 (2011-01).

[163] Ahmed R Elsherif and Mohamed M Khairy. Adaptive primary synchronization signal detection for 3GPP Long Term Evolution. In *Wireless Communications*

and Mobile Computing Conference (IWCMC), 2013 9th International, pages 1716–1721. IEEE, 2013.

[164] Leila Nasraoui, Leila Najjar Atallah, and Mohamed Siala. Robust doubly-differential primary synchronization approach for 3GPP LTE systems. In *Wireless Communications and Mobile Computing Conference (IWCMC), 2014 International*, pages 1069–1074. IEEE, 2014.

[165] Branislav M Popović and Fredrik Berggren. Primary synchronization signal in E-UTRA. In *Spread Spectrum Techniques and Applications, 2008 IEEE 10th International Symposium on*, pages 426–430. IEEE, 2008.

[166] 3GPP. LTE Feasibility study for evolved Universal Terrestrial Radio Access (UTRA) and Universal Terrestrial Radio Access Network (UTRAN). Technical report, ETSI, France, 2009. Technical Specification Name: 3GPP TR 25.912 version 9.0.0 Release 9.

[167] Giulio Bartoli, Romano Fantacci, and Dania Marabissi. AeroMACS: A new perspective for mobile airport communications and services. *IEEE Wireless Communications*, 20(6):44–50, 2013.

[168] Antonio Correas and Nikos Fistas. Aeromacs: Impact of link symmetry on network capacity. In *Integrated Communications Navigation and Surveillance (ICNS), 2016*, pages 2C4–1. IEEE, 2016.

[169] Brian Crowe. Proposed aeromacs pki specification is a model for global and national aeronautical pki deployments. In *Integrated Communications Navigation and Surveillance (ICNS), 2016*, pages 1–19. IEEE, 2016.

[170] Paolini M and Fili S. AeroMACS: A common platform for air traffic management applications. Technical report, WiMAXFORUM, 11 2016.

[171] Veeranna Daravath and Aruna Daravath. WIMAX (IEEE 802.16) broad band technology for smart grid applications. In *Communications and Signal Processing (ICCSP), 2015 International Conference on*, pages 1273–1275. IEEE, 2015.

[172] Fariba Aalamifar and Lutz Lampe. Optimized WiMAX Profile Configuration for Smart Grid Communications. *Applied Soft Computing, to be published in*, 99, 2016.

[173] D Gray. Assessing Spectrum Requirements for Smart Grid Network. Technical report, WiMAXFORUM, 7 2014.

[174] Fabio Leccese, Marco Cagnetti, and Daniele Trinca. A smart city application: A fully controlled street lighting isle based on Raspberry-Pi card, a ZigBee sensor network and WiMAX. *Sensors*, 14(12):24408–24424, 2014.

[175] WiMAXFORUM. WiMAX Forum Initiatives. http://wimaxforum.org/Page/Initiatives, 11 2016.

[176] Miguel A. del Arco-Vega, Lucas Cuadra, José Antonio Portilla-Figueras, and Sancho Salcedo-Sanz. Near-optimal user assignment in LTE mobile networks with evolutionary computing. *Transactions on Emerging Telecommunications Technologies*, pages n/a–n/a, 2016. ett.3132.

[177] Mohammad Dehghani, Kamran Arshad, and Richard MacKenzie. LTE-Advanced radio access enhancements: A survey. *Wireless Personal Communications*, 80(3):891–921, 2015.

[178] Dantong Liu, Lifeng Wang, Yue Chen, Maged Elkashlan, Kai-Kit Wong, Robert Schober, and Lajos Hanzo. User association in 5G networks: A survey and an outlook. *IEEE Communications Surveys & Tutorials*, 18(2):1018–1044, 2016.

[179] Antonis G Gotsis and Athanasios D Panagopoulos. On user association and multiple access optimisation in 5G massive MIMO empowered ultra dense networks. *Transactions on Emerging Telecommunications Technologies*, 2016.

[180] Antonis Gotsis, Stelios Stefanatos, and Angeliki Alexiou. UltraDense Networks: The New Wireless Frontier for Enabling 5G Access. *IEEE Vehicular Technology Magazine*, 11(2):71–78, 2016.

[181] Antonis G Gotsis, Stelios Stefanatos, and Angeliki Alexiou. Optimal user association for massive MIMO empowered ultra-dense wireless networks. In *2015 IEEE International Conference on Communication Workshop (ICCW)*, pages 2238–2244. IEEE, 2015.

[182] Erik G Larsson, Ove Edfors, Fredrik Tufvesson, and Thomas L Marzetta. Massive MIMO for next generation wireless systems. *IEEE Communications Magazine*, 52(2):186–195, 2014.